Doctoral Dissertation

A Study on Analysis and Utilization of Crowd-sourced Spatio-temporal Contexts from Social Media

Supervisor Professor Kazutoshi SUMIYA

Graduate School of Human Science and Environment
University of Hyogo

Shoko WAKAMIYA

2013

A Study on Analysis and Utilization of Crowd-sourced Spatio-temporal Contexts from Social Media

Shoko WAKAMIYA

Abstract

Due to the great evolution of media environment where a variety of traditional media such as videos, photos, newspapers, TV, etc., new types of massive online media such as Youtube, Niconico-Doga, Twitter, Facebook, and Foursquare have come to the forefront. Particularly, we can find a critical trend over the evolution that media consumers are no more passive receivers, rather much aggressively attending the whole process of media activity from creation to sharing often through recent popular social network sites. As for social media shared over the social network sites, the most important characteristic of social media is that they are created and shared by numerous crowds' voluntary participations reflecting their lifestyles in real world. Furthermore, compared with the conventional media, social media on the social network services may reflect various aspects of people's social activities. Thereby, the social media are the valuable source for knowing directly crowd's experiences and indirectly a variety of social phenomena as well as just logs of personal lives. For instance, user comments to video clips on video sharing websites such as YouTube and NicoNico-Doga are representing users' sentiment, opinions, interests, location information, etc. In this doctoral thesis, we describe on analysis and utilization of social media contents created by from the explosive growing social media space. Specifically, we propose our approaches not only for cooperative analysis, but also for spatio-temporal analysis of social media.

A Study on Analysis and Utilization of Crowd-sourced Spatio-temporal Contexts from Social Media

Contents

List of Tables

List of Figures

List of Publications

Journal and International Conference Papers

1. Scene Extraction for Video Clips based on the Relation of Text, Pointing Region and Temporal Duration of User Comments
   Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
   In Proc. of 3rd International Workshop on Management and Interaction with Multimodal Information Content (MIMIC 2009), pp. 289–294, Linz, Austria, September 2009
2. Towards Better TV Viewing Rates: Exploiting Crowd's Media Life Logs over Twitter for TV Rating
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2011), pp. 39:1–39:10, Seoul, Korea, February 2011
3. Discovery of Unusual Regional Social Activities using Geo-tagged Microblogs
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   World Wide Web (WWW), Special Issue on Mobile Services on the Web, Volume 14, Number 4, pp. 321–349, March 2011
4. Crowd-powered TV Viewing Rates: Measuring Relevancy between Tweets and TV Programs
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 2nd International Workshop on Social Networks and Social Media Mining on the Web (SNSMW 2011), Lecture Notes in Computer Science, 2011, Volume 6637, pp. 390–401, Hong Kong, April 2011
5. Urban Area Characterization Based on Semantics of Crowd Activities in Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of Fourth International Conference on Geospatial Semantics (GeoS 2011), Lecture Notes in Computer Science, 2011, Volume 6631, pp. 108–123, Brest, France, May 2011
6. Scene Extraction System for Video Clips using Attached Comment Interval

and Pointing Region
Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
Multimedia Tools and Applications, Volume 54, Number 1, pp. 7–25, August 2011

7. Crowd-based Urban Characterization: Extracting Crowd Behavioral Patterns in Urban Areas from Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2011), pp. 77–84, Chicago, Illinois, USA, November 2011 **[Best Student Paper Award]**
8. Crowd-sourced Urban Life Monitoring: Urban Area Characterization based Crowd Behavioral Patterns from Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 6th International Conference on Ubiquitous Information Management and Communication (ICUIMC2012), pp. 26:1–26:9, Kuala Lumpur, Malaysia, February 2012
9. Exploring Reflection of Urban Society through Cyber-Physical Crowd Behavior on Location-based Social Network
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 3rd International Workshop for Social Network and Social Web Mining (SNSM2012), pp. 168–179, Busan, Korea, April 2012
10. Urban Characteristics Extraction Based on Crowd Behavior by Tweets Analysis
    Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
    Information Processing Society of Japan (IPSJ) Transactions on Databases (TOD54), Vol.5, No.2, pp. 36–52, June 2012 (in Japanese).
11. Measuring Crowd-sourced Cognitive Distance between Urban Clusters with Twitter for Socio-cognitive Map Generation
    Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
    In Proc. of the Fourth International Conference on Emerging Databases-Technologies, Applications, and Theory (EDB 2012), pp. 181–192, Seoul, Korea, August 2012 **[Best Paper Award]**

12. Crowd-sourced Cartography: Measuring Socio-cognitive Distance for Urban Areas based on Crowd's Movement
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 4th International Workshop on Location-Based Social Networks (LBSN 2012), pp. 935–942, Pittsburgh, Pennsylvania, USA, September 2012 [Best Student Paper Award]
13. Looking into Socio-cognitive Relations between Urban Areas based on Crowd Movements Monitoring with Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
DBSJ Journal, Vol. 11, No. 2, pp. 19–24, October 2012
14. Urban Area Characterization based on Crowd Behavioral Lifelogs over Twitter
Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
Personal and Ubiquitous Computing: Volume 17, Issue 4, pp. 605–620, April 2013
15. Urban Proximity Analysis with Crowd Movements based on Location-based Social Networks
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
Information Processing Society of Japan (IPSJ) Transactions on Databases (TOD58), Vol. 6, No. 3, pp. 159–176, June 2013 (in Japanese)
16. Geo-social Neighborhood Area Search based on Urban District Proximity using Crowd Experience over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
DBSJ Journal, Vol. 12, No. 1, pp. 97–102, June 2013 (in Japanese)

Technical Reports

1. A Shared Video Viewing System using User Comments based on Temporal Duration and Pointing Region
Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
In Proc. of Information Processing Society of Japan (IPSJ) Kansai-Branch Convention, pp. 211–214, October 2008 (in Japanese)
2. Scene Extraction for Shared Video based on the Relation of Text, Pointing

Region and Temporal Duration of User Comments
Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
In Proc. of the 1st Forum on Data Engineering and Information Management (DEIM Forum 2009), E9-3, March 2009

3. Scene Retrieval Method for Shared Video using Attached User Comments and Web Page Context
Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
In Proc. of the 2nd Forum on Data Engineering and Information Management (DEIM Forum 2010), F1-5, March 2010 (in Japanese)
4. Scene Extraction Method to Complement Web Page using User Comments Attached to Shared Video
Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
In Proc. of Information Processing Society of Japan (IPSJ) 50th Anniversary, 72nd National Convention of IPSJ, pp. 775–776, March 2010 (in Japanese)
5. Video Scene Retrieval Method Using Attached User Comments Based on Automatic Query Expansion
Shoko Wakamiya and Kazutoshi Sumiya
In Proc. the 2nd International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2010), pp. 23–29, August 2010
6. Twitter-based TV Audience Behavior Estimation for Better TV Ratings
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 3rd Forum on Data Engineering and Information Management (DEIM Forum 2011), A1-4, February 2011
7. Urban Area Characterization by Analyzing Crowd Activities in Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 2nd Symposium on Social Computing (SoC2011), pp. 27, June 2011
8. Crowd-based Image of City: Urban Area Characterization through Geographic Regularity over Twitter

Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 3rd International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2011), pp. 128–137, August 2011

9. Crowd Classification based on Behavioral Patterns over Location-based Social Networks
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 4th Forum on Data Engineering and Information Management (DEIM Forum 2012), D8-3, March 2012 (in Japanese)

10. Social Cognitive Map Generation based on Crowd's Movement over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 3rd Symposium on Social Computing (SoC2012), pp. 49–54, June 2012

11. Socio-Cognitive Urban Cartography with SNS-based Crowd's Movement
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 4th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2012), August 2012

12. Geo-social Neighborhood Area Search based on Urban District Proximity using Crowd Experience over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 5th Forum on Data Engineering and Information Management (DEIM Forum 2013), A3-3, March 2013 (in Japanese) [Student Presentation Award]

13. Analyzing Distortion of Geo-social Proximity using Massive Crowd Moving Logs over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 4th Symposium on Social Computing (SoC2013), pp. 23–28, June 2013

Article

1. Fusion of Image and Heterogeneous Media
   Kazutoshi Sumiya, Daisuke Kitayama, and Shoko Wakamiya
   The Journal of the Institute of Image Information and Television Engineers, Vol. 66, No. 11, pp. 896–902, November 2012 (in Japanese) [Best Author]

Submitting

1. Exploring Geospatial Cognition based on Location-based Social Network Sites
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   World Wide Web (WWW), Special Issue on Internet and Web Information Systems, 2013
2. Exploring Latent Relations between Urban Areas and Crowd Behavior from Location-based Social Networks for Urban Area Characterization
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   Journal of Database Management, 2013
3. Crowd Footprints Network for Social-Urban Analytics with Location-based Social Networks
   Ryong Lee, Shoko Wakamiya, Kazutoshi Sumiya
   Distributed and Parallel Databases, Special issue on Data Management and Analysis in Location-Based Social Networks, 2013
4. Fusion of Image and Heterogeneous Multimedia
   Daisuke Kitayama, Shoko Wakamiya, Yuanyuan Wang, and Kazutoshi Sumiya
   Image Laboratory, 2013 (in Japanese)

Chapter 1 Introduction

1.1 Motivation and Our Approaches

We have witnessed the great evolution of media environment where a variety of traditional media such as videos, photos, newspapers, TV, etc. are transformed into new types of massive on-line media such as YouTube [1], NicoNico-Doga [2], Twitter [3], Facebook [4], and Foursquare [5]. Characteristically, we can find a critical trend over the evolution that media consumers are no more passive receivers, rather much aggressively attending the whole process of media activity from creation to sharing often through recent popular social network sites. As for social media shared over the social network sites, the most important characteristic is that they have been created and shared by numerous crowds' voluntary participations reflecting their real-world lives. Furthermore, compared to the conventional media, social media may reflect various aspects of people's social activities. Thereby, the social media are not just storage and transmission channels or tools used to store and deliver data or information, but the valuable source for grasping directly crowd's viewpoints and experiences and indirectly a variety of social phenomena via the media. For instance, user comments to video clips on video sharing websites such as YouTube and NicoNico-Doga are representing various users' viewpoints, sentiments, opinions, interests, etc. In addition, microblogs with location information over microblogging sites such as Twitter are reflecting the real space via crowd behavior and activities. Therefore, we can take advantages of the explosive growing social media, we will be able to conduct various analyses. In this doctoral dissertation, we conduct analysis and utilization by focusing on two important characteristics of social media, that is, a) collaborative analysis and b) spatio-temporal analysis as shown in Figure 1.

Cooperative Analysis :

Scene Extraction for Shared Video Clips We propose a system to easily retrieve video scenes relevant to their interests. The system analyzes both text and non-text aspects of a user's comment and then retrieves and displays relevant scenes along with attached comments.

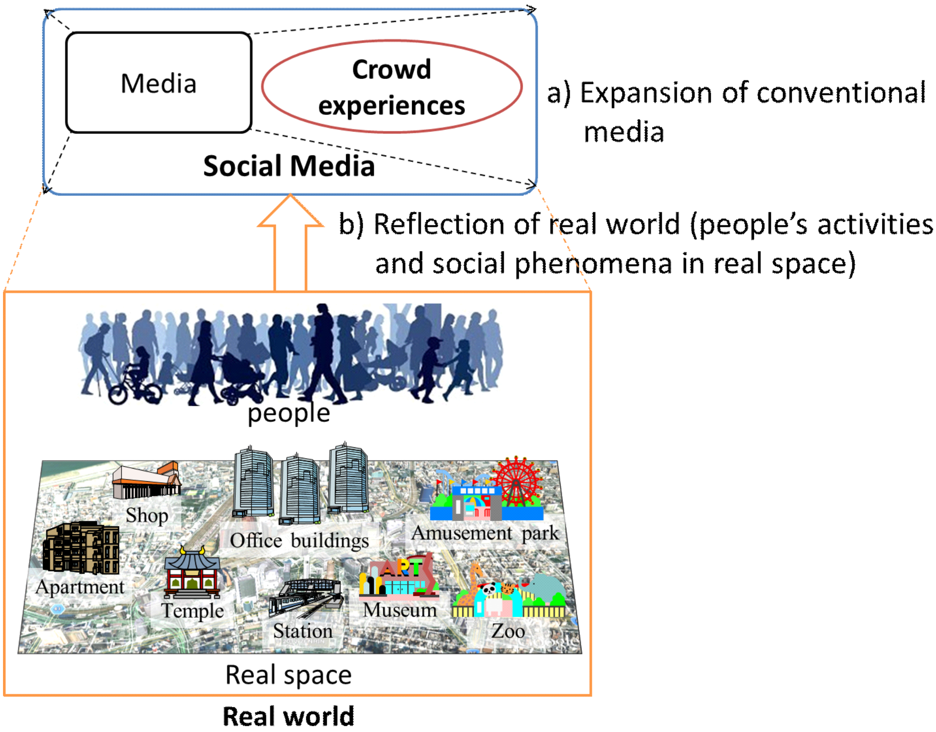


Figure 1: Concept of Social Media

The text analysis works in tandem with non-text features, namely, the selected area and temporal duration associated with user comments. In this way, our system supports a better-organized retrieval of scenes that have been commented on with a higher degree of relevancy than conventional methods, such as using matching keywords.

Crowd-powered Video Rating We propose a method to measure popularity of video clips broadcast through several video media in the light of the evolving TV lifestyles beyond home environments and the increasing video media. Particularly, we propose a video rating method by exploiting Twitter where we can easily find crowd voices relative to video watching.

Spatio-temporal Analysis :

Crowd-sourced Urban Area Characterization We challenge to extract urban characteristics by observing crowd behavior in urban areas by utilizing crowd microblogs over the social networking sites. For this, we first present a model to deal with crowd behavioral logs on the social network sites as a representing feature of urban space's characteristics, which will be used to conduct crowd-based urban area characterization. Based on this crowd behavioral feature, we will extract significant crowd behavioral patterns in a period of time.

Crowd-sourced Cartography We attempt to exploit massive microblogs with location information for generating a socio-cognitive map, whose purpose is to deliver much simplified and intuitive perspective of urban space in the form of a map. For the purpose, we will measure socio-cognitive distance among urban clusters based on human mobility to represent accessibility of urban areas based on crowd's movements. Finally, we generate a socio-cognitive map by measuring influential strengths of urban clusters.

1.2 Outline of the Doctoral Thesis

In this thesis, we introduce our approaches on analysis and utilization of crowd-sourced spatio-temporal contexts from social media. This thesis consists of six chapters, including this chapter as the introduction.

In Chapter 2, in order to position our research comparing with others and show the value of our research, we introduce related work.

In Chapter 3 and 4, we describe our approaches for utilizing social media as enhanced media comparing with conventional ones by people's participation and their various logs. Concretely, we measure media relation and popularity with crowd experiences; i) crowd-sourced video scene summarization and ii) crowd-powered TV viewing rates. As for a), we attempted to extract relevant scenes by measuring relations among user comments attached with pointing regions and time intervals. In the case of ii), the interaction from people to media like TV programs shows significant changes. Therefore, the way measuring the value of media is also changing. For this, we challenge to rate TV programs by collecting TV watching logs (experience) from crowd lifelogs over Twitter. Because the site was not designed for this specific goal to collect the TV-related Twitter messages, so-called tweets, we need to identify those that are relevant to TV programs.

In Chapter 5, we show a method to extract crowd-centered urban characteristics by observing crowd activities. Conventionally, in order to grasp characteristics of urban areas, a questionnaire-based survey has been conducted. However, it is time-consuming and hard to know dynamic characteristics. Therefore, we proposed a method to dynamically and lightly crowd-sourced urban characteristics by extracting latent patterns of crowd behavior in urban areas and classify the urban areas. In Chapter 6, we explain a method to measure crowd-based proximity among urban areas by monitoring crowd movings. Here, we focus on the difference between geographic proximity and cognitive proximity which is deformed by crowd lifestyles depending on transportation infrastructure. Therefore, we measure proximity between urban areas in terms of spatial, temporal, and social aspects and extract vicinity areas.

Finally, we discuss future directions of my research in Chapter 7 and conclude this doctoral thesis in Chapter 8.

Chapter 2 Related Work

2.1 Cooperative Analysis

Most of the research concerning video sharing websites has focused on video retrieval in these sites. Various approaches such as content-based video retrieval [6, 7] and video retrieval by analysis of viewing history [8] have been described. In our proposed method, however, we focused on exploiting user comments attached to a video clip for extracting desired scenes.

There are several related studies using user comments associated with particular scenes of a video clip. Yamamoto et al. [9] developed an on-line video-annotation system called iVAS that let users associate detailed annotations such as text annotation, personal impressions, and evaluations with scenes on a video. They also developed a video-sharing system called Synvie [10, 11] that extended iVAS to automatically extract deep-content-related information about video content as annotations based on social activities, especially user comments and weblog authoring, associated with the content of video clips on the Web. They proposed a scene retrieval system based on scene tags generated automatically from annotations. The focus of their research is on extracting annotations not only from scene comments but also from weblog entries quoting the video scenes and using the extracted information with various applications. Meanwhile, our focus is on using the relation of user comments with the pointing region and temporal duration to retrieve scenes.

Kitayama et al. [12] proposed a method that generates comment sets using the pointing region and temporal duration for the purpose of organizing large numbers of comments in video sharing systems. Kimura et al. [13] developed a video editing support system that uses the gaze of the user while watching a video clip. In this system, user gaze represents the user's viewing of the pointing region and temporal duration. Therefore, these studies utilize only non-text information gathered from user comments or the user's gaze. However, we need to consider text information because it would be difficult to adequately determine the relation of user comments without it. Uehara et al. [14] proposed a method to automate the annotation of TV programs based on information from

Internet bulletin boards. Miyamori et al. [15] proposed a method of generating various views of broadcast content using viewers' perspectives expressed in live chats on the Web, such as chronological and ranking views reflecting viewers' perspectives, views reflecting the perspectives of all viewers and particular viewers, and views reflecting the intensities of responses and the degree of emotions like enjoyment and disappointment. Although these studies targeted only TV programs, not video clips, they showed the effectiveness of user reviews for TV programs, such as Internet bulletin boards and live chats that are similar to user comments for video clips. So our method can be improved by using the analysis ways of unstructured texts.

There are several studies using text extracted from video or professionally written to describe the video content. Miura et al. [16] used closed captions to extract a principal video object that appeared in each video shot of a TV program. Because closed captions are based on the speech of the people in the programs as well as on narration relevant to the scenes being broadcast, they have much information about scenes and are structured. However, our method utilizes users' comments, which are freely written when users are viewing a video clip. Natural language analysis is not efficient for users' comments because they are often informal and do not use proper grammar. Dao et al. [17] proposed a method of using non-ambiguous temporal pattern mining based on TPrefixSpan algorithm [18] that was proposed to clarify ambiguous representation of Allen's model [19] and web-casting text to detect specific events in sports videos. Fukino et al. [20] proposed a method to generate summaries of football videos by using a set of glue operations [21] for each event that was detected from a news article. The texts they used were summaries generated by people. Therefore, these texts differ from both closed captions and users' comments that are synchronized with the video. Additionally, their method covered only sports videos, which have consistent scene structures. Our method can be applied not only to sports but also to other types of video clips. Wu et al. [22] proposed a method for detecting and retrieving videos of the same scene (scene duplicates) from broadcast video archives. A scene duplicate is composed of different pieces of footage of the same scene, the same event, at the same time, but filmed from different viewpoints.

They focused on object motion in videos and devised a video matching approach based on the temporal pattern of discontinuities obtained from feature point trajectories. In our method, however, we extract scenes of the same object and event from a video clip utilizing only user comments without video analysis. Fukuhara et al. [23] developed a system for collecting and analyzing blog articles to gain an understanding of the concerns of people from collective and personal viewpoints. Their approach analyzes relationships between blog articles and real temporal data, extracts a topic of interest, and identifies trends. Glance et al. [24] developed a system called BlogPulse, which extracts trends from collected blog articles. Using keyword occurrence rates over a given period of time, the system classifies current trends.

Furthermore, recently, microblogging services represented by Twitter grow more popular and have been aggregating lots of researchers' attentions as a critical research topic in various fields. As initial work, the usage and role of Twitter in creating a social community on the basis of its basic functions were examined by Java et al. [25], Zhao et al. [26], Krishnamurthy et al. [27], and Cha et al. [28]. In these studies, Twitter was investigated for its social networking role, that is, how it would be used to send massive amounts of short messages about social activities. Obviously, even major global news channels refer to Twitter as an important social channel, and many people are aware of its role as an uncontrolled and uncensored communication channel. In addition, several research studies focused on the role of tweets as a novel media to represent crowd opinions. O'Connor et al. [29] compared the measures of public opinion from polls with ones from the analysis of tweets. Diakopoulos et al. [30] demonstrated an analytical methodology including visual representations and metrics that aid in making sense of the sentiment of social media messages around a televised political debate. By finding tweets relative to TV watching, we measure popularity for several video media. Sawai et al. [31] have proposed a method to recommend TV programs based on relations among users over social networking.

In order to attempt to measure TV ratings based on Twitter by looking into Twitter messages including TV-related words made of TV program titles or some verbs such as ‘view’ or ‘watch’ and ‘TV’, we examined the usefulness of Twitter as a means to carry out a poll [32, 33].

2.2 Spatio-temporal Analysis

In recent years, social networking sites can be regarded as a novel source, where we are able to easily monitor a great deal of daily crowd lifelogs. Obviously, this kind of unprecedented popularity by numerous people around the world is reflecting drastic changes towards reciprocal benefits among people through sharing personal lifelogs. While there would still be seemingly endless concerns about privacy related to personal location sharing, it must be a critical issue more and more to exploit the crowd-sourced data in the academic field as well as the business world in terms of social and individual benefits from the new open sharing space.

As for location identification or labeling to user-written messages over Twitter, on behalf of prevalence of location sensing devices including recent GPS-embedded smartphones, several granularity levels of location can be automatically attached into the user messages from the fine latitude/longitude coordinates to the city name as a coarse representation. In case of a study conducted by Cheng et al. [34] for the purpose of getting more geographically related tweets, they attempted to reveal users’ location information only from the written text by referring to other geo-tagged tweets with fine location coordinates. On the other hand, according to the survey by Sysomos [35] in 2010, there are already 73% users who have submitted geo-tagged tweets compared to 44% in 2009. While we still have to be aware of the privacy concern in exploring individual data, Barkhuus et al. [36] exemplified the usefulness of open sharing of individual updates including locations through mobile devices for a group of people in a field study, where participants can unexpectedly take advantages of awaring friends’ updates less interrupting each other’s daily activities. Furthermore, Hightower [37] discussed the importance of semantic location labeling by taking into accounts the activities of people on the places.

For exploiting the further useful use cases, lots of research work have been studied. Zheng et al. [38] presented a method to utilize the social network as a location information search framework by extracting knowledge over the location data based on GPS and users' comments at various locations to answer two typical socio-geographic questions: If we want to do something, where shall we go? If we visit some place, what can we do there? By modeling a location-activity relation into a matrix, the former question is answered by activity recommendation to given a location query, and the latter one is also resolved by location recommendation given an activity query. In this work, authors focused on determining the correlation between locations and crowd activities, while we are distinctively focusing on regular crowd activities on locations.

Interestingly, some researches focusing on cooperation with Twitter for analyzing some natural incidents in the real world have been introduced. De Longueville et al. [39] analyzed the temporal, spatial and social dynamics of Twitter activity during a major forest fire incident in the South of France in July 2009. Sakaki et al. [40] constructed an earthquake reporting system in Japan using tweets which are posted from each Twitter user regarded as a sensor. In this method, tweets which are reporting the occurrence of earthquakes are extracted by using textual information. Cheng et al. [41] studied human mobility patterns revealed by the check-ins over location sharing services and explored the corresponding factors that influence mobility patterns, in terms of social status, sentiment, and geographic constraints. This study focused on analysis of massive personal trace data based on characterizing geographic facilities. In our previous work [42, 43], we also proposed a method to discover local social and natural events by monitoring unusual statuses of local users' activities utilizing geo-tagged crowd lifelogs over Twitter.

Due to the fast urbanization in the modern age, it is not easy to draw images of urban areas in mind, especially, to unfamiliar cities. For this problem, lots of interdisciplinary studies computationally to investigate urban characterization have been proposed. In our previous work [44, 45, 46, 43], we proposed a method to characterize urban areas by detecting significant latent patterns of crowd's behavior exploiting geo-tagged tweets extractable from Twitter. Yuan et al. [47] proposed a framework for discovering different functional regions such as educational areas, entertainment areas, and regions of historic interests in a city using both human mobility based on taxies trajectory and POIs. Kurashima et al. [48] developed a system for browsing actual experiences related to a specific location and time period extracted by means of association rules from blog articles. Advanced from our previous work, we focus on measuring proximities cognitively recognized among urban areas based on massive crowd's experiences through location-based social network sites for generating a socio-cognitive map which can intuitively represent complicated urban structure.

In order to represent specific information with maps, thematic maps are required rather than general reference maps. Therefore, lots of technologies for cartography satisfying various purposes have been studied recently. Grabler et al. [49] presented a method for automatically generating destination maps to navigate to a given location from anywhere in a given area of interest. The system decided map elements using both vision-based image analysis and web-based information extraction techniques. This method aimed to create a navigational map in much simplified and personalized way. In addition, there are interesting and cognitive maps like cartogram which can represent location-based statistical information by deforming area or distance. Shu et al. [50] developed a system to generate animated map by means of interactions with a user and the system. Mislove et al. [51] investigated mood throughout a day in the U.S. by examining Twitter and visualized hourly deformed maps in terms of Twitter users' mood observable from their textual messages. In our proposed method, we create a socio-cognitive map by computing an influential strength of an urban cluster and relations among urban clusters based on crowd's mobility.

Chapter 3 Scene Extraction for Shared Video Clips using Attached Comment Interval and Pointing Region

3.1 Introduction

Many video clips are shared on-line on video sharing websites. Users can write comments about an entire video clip on various sites such as YouTube [1] and Google Video [52]. However, on sites such as Japan's Nico Nico Douga [2], users can write comments about particular scenes, and these comments are synchronized with the playing of the video clip and displayed on the screen for a definite time interval. Various comment-posting systems or systems that enable video annotation by users have been developed [53, 9, 54].

Because any user can easily write comments on these videos, one user's comments may be irrelevant to another user. On the other hand, comments that are relevant to the user can be used to retrieve scenes of the user's interests. For example, when a user is interested in a batter and views a video clip of a baseball game, the user wants to see more scenes related to him. Scenes in which he appears are likely to have user comments related to him. When a scene has many user comments about the batter, these scenes can be regarded as being related to that user's interest. We used this idea to develop a method for extracting scenes related to the user's interests. Video analysis techniques such as image processing and speech recognition are useful for recognizing objects in a video clip. However, applying these techniques is expensive and it would not be possible to analyze all shared video clips. Therefore, we focus on utilizing user comments attached to a video clip to grasp the contents of the video clip.

We developed a method of determining relations among user comments on a video clip and then extracting scenes related to a user's interests on the basis of that relation. Matching keywords is the simplest method of using user comments, but is not adequate for extracting relevant scenes because the relation among such comments is not clear. Therefore, we developed a posting system where users can write comments with a specified pointing region and temporal duration (see Figure 2). In this system, a comment consists not only of text, but also of the specified pointing region and its specified temporal duration. These user comments enable us to extract scenes relevant to the user's interests at lower cost than video analysis. We can also extract the relations among user comments. Figure 3 is an example of the user comments in our system. The x and y axes are screen coordinates, and t is the time axis. The pointing region is a rectangular area on the video screen selected by the commenter and consists of coordinates and area size. The temporal duration that is represented by the arrow in Figure 3 is the time interval specified by the user to display the comment and its pointing region and consists of a starting point and an ending point. The text is treated as a keyword set consisting of nouns, verbs, adjectives, and/or adverbs. In our system we divide a video clip into scenes based on video analysis techniques such as image processing [55, 56, 57] and speech recognition [58, 59].

In our proposed method, users can view scenes and their attached comments about the specific object and event they are interested in merely by selecting a comment of interest. Take, for example, a user watching a baseball game who would like to see a batter as the object and getting a bat as the event. The object and event are defined based on the comment's pointing region and temporal duration, as well as the comment's text. In our system, as a user selects a comment entered by previous users, we determine the object using the selected comment's text, the pointing region, and other comments. Then, we search for the event using the selected comment's text, the temporal duration, and other comments. Our method has several strengths.

1. **The advantage of the object determination method** By using the pointing region, our method can locate scenes relevant to the user's comment even when the comment does not contain keywords found in other comments about the same object. Also, our system can distinguish where keywords common to several comments in fact refer to different scenes.
2. **The advantage of the event determination method** Since video clips are rarely annotated by the uploader, background information concerning particular events appearing in the video clip is not available. Even so, our

method can locate these events because comments are linked with temporal duration.

In addition, this method based on user comments with a specified pointing region and temporal duration can be used with almost any video clips recorded by professional or semiprofessional cameramen using a fixed camera or conventional camera operation. It can be effective for various genres, such as sports, news, documentaries, concerts, festivals, and shows.

Figure 2: Screenshot of Comment Posting System

3.2 Object Determination based on Pointing Region

3.2.1 Concept of Object Determination

When a user enters a comment on the screen, the user indicates the object that is the focus of the comment by the pointing region. In other words, an object is defined by keywords in the comment and by the pointing region (see Figure 4). As a user selects a comment entered by previous users in order to search scenes of an object, we determine whether an object indicated by the selected comment matches other comments comparing keywords and spatial relations of the

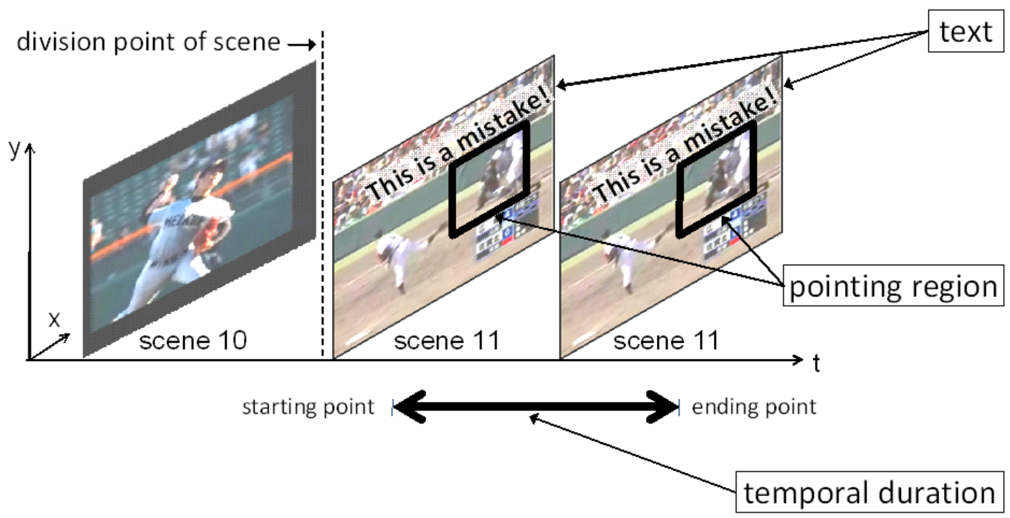


Figure 3: User Comment with a Specified Pointing Region and Temporal Duration

pointing regions such as their degree of overlap or relative positional relation.

$$C_{object} = \{c_o | Cov(c_s, c_o) \geq \alpha \; or \; Pos(c_s, c_o) \geq \beta\} \tag{1}$$

c_s is a user-selected comment and c_o is an extracted comment indicating the same object as c_s. The function Cov calculates the degree of overlap of the pointing regions and the similarity of the keywords, and the function Pos calculates the relative positional relation of the pointing regions and the similarity of keywords. If Cov or Pos is higher than the threshold, we identify c_o as a comment related to the same target object.

With this method, we can extract a comment as indicating the same object as the selected comment even when the comment does not contain keywords found in other comments associated with the same object. We can also identify when a comment indicates a different object than the selected comment even when they contain the same keywords.

3.2.2 Object Determination using Degree of Overlap of Pointing Regions

In a video clip with still or steady images, the same object in separate scenes tends to appear at near absolute positions. For example, in two scenes from a video clip of a football game in Figure 5, the same keeper appears on the left

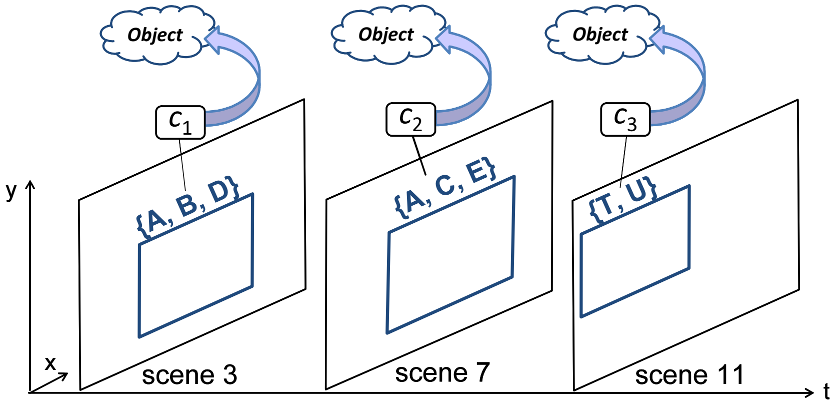


Figure 4: Concept Image of Object

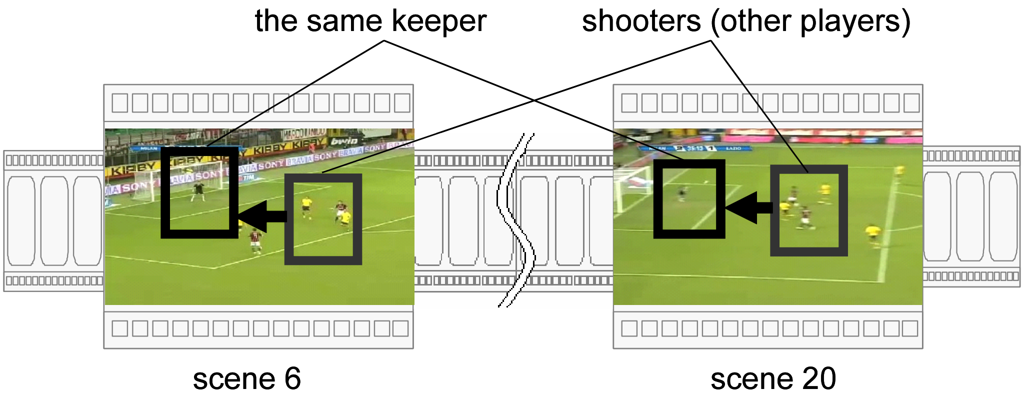


Figure 5: Motivating Example of Object Determination

side of the screen and is specified by the comments' pointing region. So, the degree of overlap of the comments' pointing regions will tell us if the objects they refer to are the same.

We determine whether objects in separate scenes are the same by calculating the function Cov as follows.

$$Cov(c_s, c_i) = CovR(c_s, c_i) \times (CovK(c_s, c_i) + \gamma) \tag{2}$$

$$CovR(c_s, c_i) = \frac{|R(c_s) \cap R(c_i)|}{|R(c_s) \cup R(c_i)|} \tag{3}$$

$$CovK(c_s, c_i) = \frac{|K(c_s) \cap K(c_i)|}{|min(K(c_s), K(c_i))|} \tag{4}$$

where c_s is a user-selected comment, and c_i is one of other comments. The

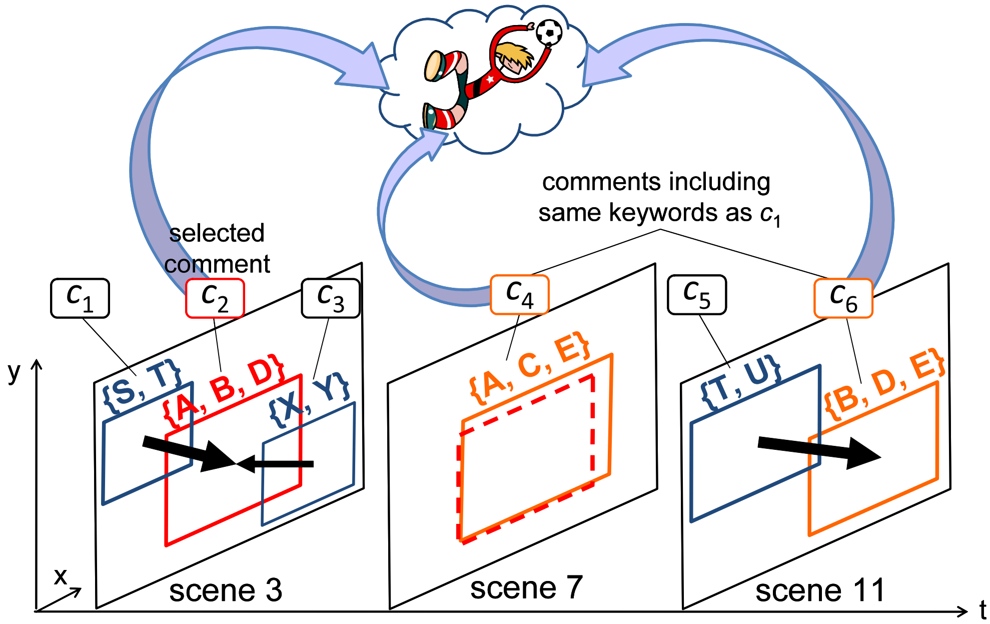


Figure 6: An Example of Object Determination using Spatial Relations of Pointing Regions

function $CovR$ calculates the degree of overlap of the pointing regions and the function R denotes the pointing region's information such as the coordinates and area size. The function $CovK$ calculates the similarity of keywords. The function K extracts keywords from a user comment and the function min extracts the minimum value by comparing with the numbers of keywords of two user comments, and γ is a constant that prevents $CovK$ from becoming 0.

In Figure 6, when the user selects c_2 in scene 3, we extract c_4 in scene 7 as indicating the same object as the selected comment. That is because the degree of the overlap of c_2's pointing region represented by the dotted line in scene 7 and c_4's pointing region and the similarity of keywords for c_2 and c_4 are high.

3.2.3 Object Determination using Relative Positional Relation of Pointing Regions

In scenes in which the same objects appear, their relative positional relation tends to be the same. For example, in both scenes from a video clip of a football game in Figure 5, the same keeper appears on the other players' left. This occurs frequently due to the 180-degree rule that is one of basics of film making [60, 61, 62]. This rule states that two or more characters in the same scene should always have the same left or right relationship to each other. It also ensures that the relative positions of the objects on the screen in a scene are unchanged. Therefore, the relative positional relation of the pointing regions of comments in the same scene will indicate if the objects they refer to are the same.

We determine whether objects in separate scenes are the same by calculating the function Pos as follows.

$$Pos(c_s, c_c, c_i, c_j) = RelP(c_s, c_c, c_i, c_j) \times CovK(c_s, c_i) \times CovK(c_c, c_j) \tag{5}$$

$$RelP(c_s, c_c, c_i, c_j) = \begin{cases} 1 \; (P(c_s, c_i) = P(c_c, c_j)) \\ 0 \; (P(c_s, c_i) \neq P(c_c, c_j)) \end{cases} \tag{6}$$

where c_c and c_j are comments that are linked with c_s and c_i in another scene. The function P calculates the relative position of the pointing regions by comparing whether the pointing region's position is upper and more toward the left than the other one in the same scene. The function $RelP$ returns 1 when the positional relationship of the relative positions of comment pairs is the same.

In Figure 6, we extract the relative positional relations of the pointing regions of the selected comment, c_2, and related comments, c_1 and c_3, from scene 3. We also extract a comment that has the same keywords as the selected comment, c_6, and another related comment, c_5, from scene 11. In this case, the selected comment is located to the lower right of c_1 and to the upper left of c_3, and c_6 is located to the lower right of c_5 in scene 11. Here, a lower right of the pointing regions is determined for c_1 and c_5. So, we extract c_6 as referring to the same object as the selected comment.

3.3 Event Determination based on Temporal Duration

When sequential events in separate scenes are the same, the temporal relation between each sequential event is the same. Also, in many cases different users use the same keywords to comment on an event. In Figure 7, for example, the same sequential event, i.e., scoring with a header, happens in both scene 15 and scene 30. The event in each of these two scenes consists of the header and the

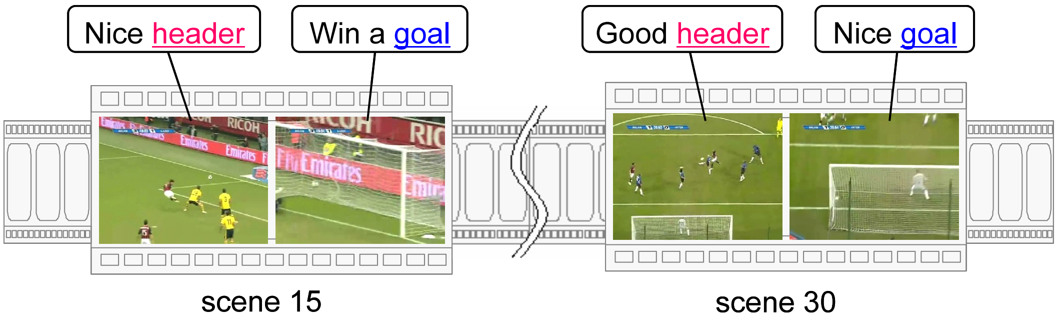


Figure 7: Motivating Example of Event Determination

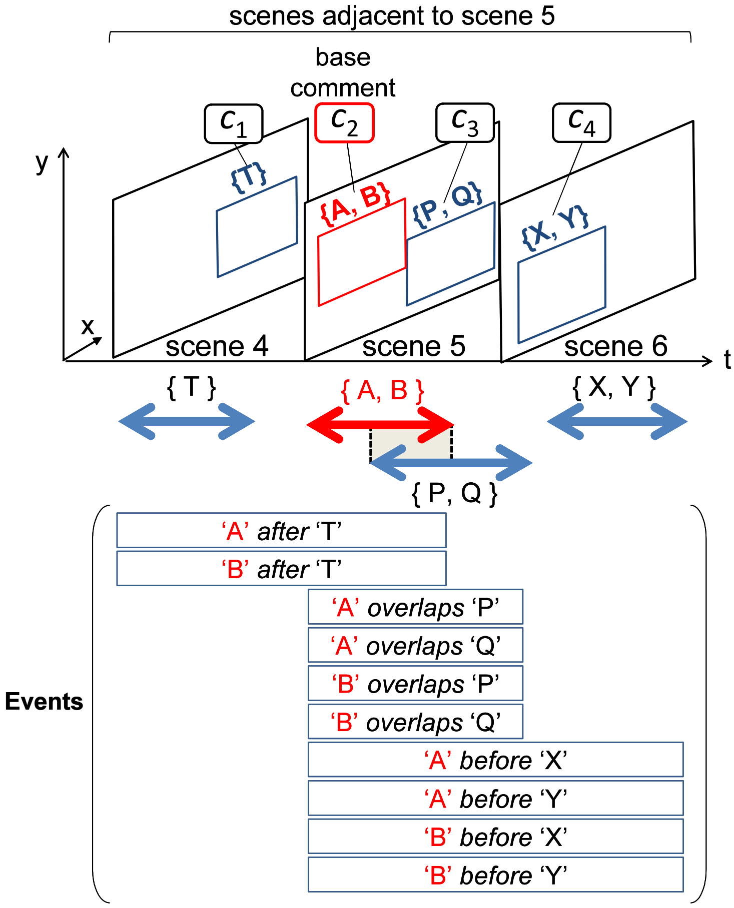


Figure 8: Concept Image of Event

score, in that order. Users comment on each event using the same keywords, such as 'header' and 'goal'. So we define an event as indicated by common keywords in comments attached to adjacent scenes (see Figure 8).

We calculate as follows to determine whether events in separate scenes are the same.

$$C_{event} = \{c_e | RelT(c_s, c_e, c_{s_p}, c_{e_p}) \times PairK(c_s, c_e, c_{s_p}, c_{e_p}) = 1\} \quad (7)$$

where c_s is a user-selected comment and c_e is an extracted comment indicating the same event as c_s. c_{s_p} is a comment that was attached to the scene of c_s, and c_{e_p} is a comment that was attached to the scene of c_e. If the function $RelT$ and $PairK$ returns 1, we detect c_e as a related comment referring to the same event.

$$RelT(c_s, c_{s_p}, c_i, c_{i_p}) = \begin{cases} 1\ (T(c_s, c_{s_p}) = T(c_i, c_{i_p})) \\ 0\ (T(c_s, c_{s_p}) \neq T(c_i, c_{i_p})) \end{cases} \quad (8)$$

$$PairK(c_s, c_i, c_{s_p}, c_{i_p}) = \begin{cases} 1\ (k_s = k_i\ and\ k_{s_p} = k_{i_p}) \\ 0\ (k_s \neq k_i\ or\ k_{s_p} \neq k_{i_p}) \end{cases} \quad (9)$$

The function T denotes the relation between the temporal duration of a comment pair. The relation is defined using the relation types based on Allen's time interval model [63] (see Table 1). In Table 1, t_{s_s} is the starting point of a selected comment and t_{s_e} is its ending point. t_{i_s} is the starting point of another comment and t_{i_e} is its ending point. The function $RelT$ is 1 when the relationship of the temporal durations of comment pairs is the same. k_s, k_i, k_{s_p} and k_{i_p} are each keyword in c_s, c_i, c_{s_p} and c_{i_p}. The function $PairK$ is 1 when the two keywords in a pair are the same.

The system searches for comments with the same keywords and temporal duration and determines whether these comments refer to the same event. In Figure 9, when a user selects c_1 in scene 5, we extract the relation of the temporal duration between the selected comment and related comments, c_2 and c_3, in the scenes adjacent to scene 5, c_4 in scene 13 because this comment contains the same keyword 'A' and related comments, c_5 and c_6, in the scenes adjacent to scene 13. Then, the keywords extracted from the comments are paired with the keywords in the selected comment, making eight different pairs. The keywords extracted from the comments are also paired with the keywords in the comment containing the same keyword as in the selected comment, also making eight different pairs. If both the two keywords in the selected comment's

Table 1: The Relation of Temporal Duration

Relation type	Determination condition	Order
before	$\lvert t_{s_e} - t_{i_s}\rvert > 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
after	$\lvert t_{i_e} - t_{s_s}\rvert > 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
equal	$\lvert t_{s_s} - t_{i_s}\rvert = 0$ and $\lvert t_{s_e} - t_{i_e}\rvert = 0$	$t_{s_s} = t_{i_s}$ and $t_{s_e} = t_{i_e}$
meets	$\lvert t_{s_e} - t_{i_s}\rvert = 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
met-by	$\lvert t_{i_e} - t_{s_s}\rvert = 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
overlaps	$\lvert t_{s_e} - t_{i_s}\rvert > 0$ and $t_{s_e} > t_{i_s}$	$t_{s_s} < t_{i_s}$ and $t_{s_e} < t_{i_e}$
overlapped-by	$\lvert t_{i_e} - t_{s_s}\rvert > 0$ and $t_{i_e} > t_{s_s}$	$t_{s_s} > t_{i_s}$ and $t_{s_e} > t_{i_e}$
during	$\lvert t_{s_s} - t_{i_s}\rvert > 0$ and $\lvert t_{s_e} - t_{i_e}\rvert > 0$	$t_{s_s} > t_{i_s}$ and $t_{s_e} < t_{i_e}$
contains	$\lvert t_{s_s} - t_{i_s}\rvert > 0$ and $\lvert t_{s_e} - t_{i_e}\rvert > 0$	$t_{s_s} < t_{i_s}$ and $t_{s_e} > t_{i_e}$
starts	$\lvert t_{s_s} - t_{i_s}\rvert = 0$ and $\lvert t_{s_e} - t_{i_e}\rvert > 0$	$t_{s_s} = t_{i_s}$
finishes	$\lvert t_{s_e} - t_{i_e}\rvert = 0$ and $\lvert t_{s_s} - t_{i_s}\rvert > 0$	$t_{s_e} = t_{i_e}$

pair and the two keywords in the comment's pair in other scenes are the same, this comment pair is identified as referring to the same event as the selected comment's pair. In this case, because the temporal duration's relationship of the pair 'A' and 'Q' in the selected comment's pair and in the same keyword comment's pair is the same *overlaps*, we consider these comment pairs to be associated with the same event.

3.4 Scene Extraction

3.4.1 Types of Scene Relations

We define four types of scene relations on the basis of the objects and events the comments are about (see Table 2). Other comments utilized in identifying the event, such as the paired comments of the selected comment and the paired comments of the comment including the same keyword as the selected comment, are included in the calculation, and the system determines whether these comments refer to the same object. We describe the identification result of object indicated by these comments as 'Paired-object', and the identification result of object indicated by the selected comment and the comment including the same

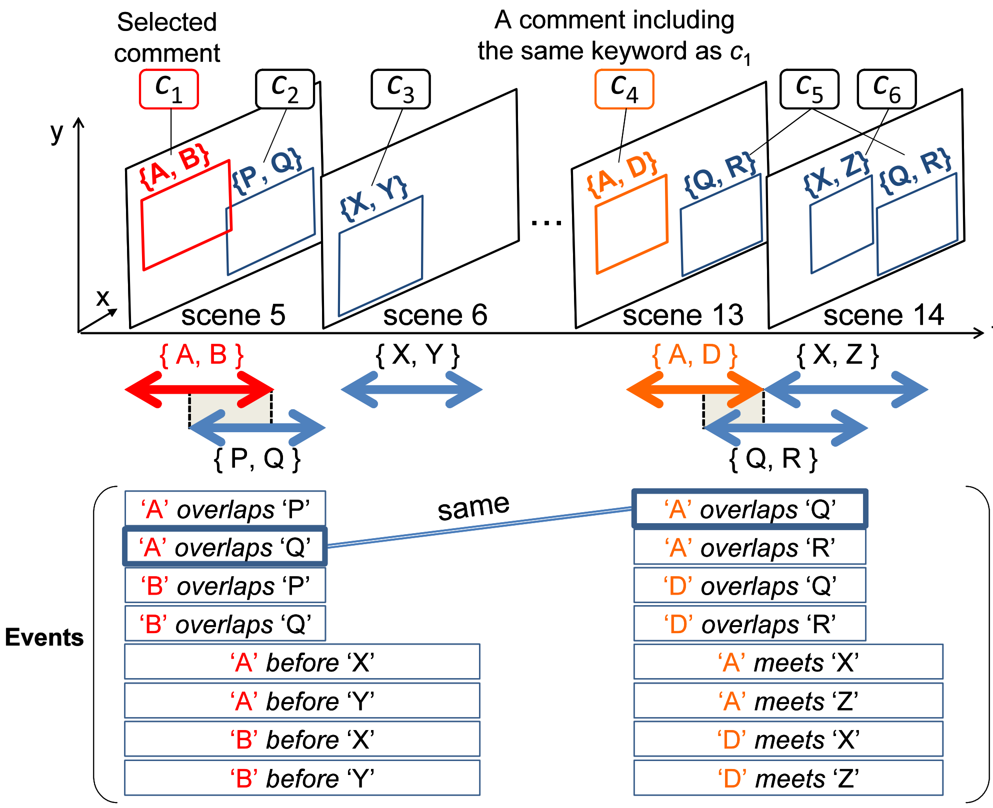


Figure 9: An Example of Event Determination using Temporal Durations

keyword as 'Selected-object' in Figure 10 and Table 2. The system uses these to determine the type of relationship between the scene of the comment that a user is interested in and of the other comments.

EQUAL

When comments share the object and event.

OBJECT

When comments share both the selected-object and event and paired-object indicates a different object, or when the comments share the paired-object only and selected-object indicates the same object or a different object.

EVENT

When comments share the event only, and paired-object indicates the same object or a different object, or when neither the selected-object nor event is shared and paired-object indicates the same object.

NO RELATION

Table 2: Relation Types between Scenes

		Objects			
	Selected-object	same	same	different	different
	Paired-object	same	different	same	different
Events	same	*EQUAL*	*OBJECT*	*EVENT*	*EVENT*
	different	*OBJECT*	*OBJECT*	*EVENT*	*NO RELATION*

When neither the object nor event is shared.

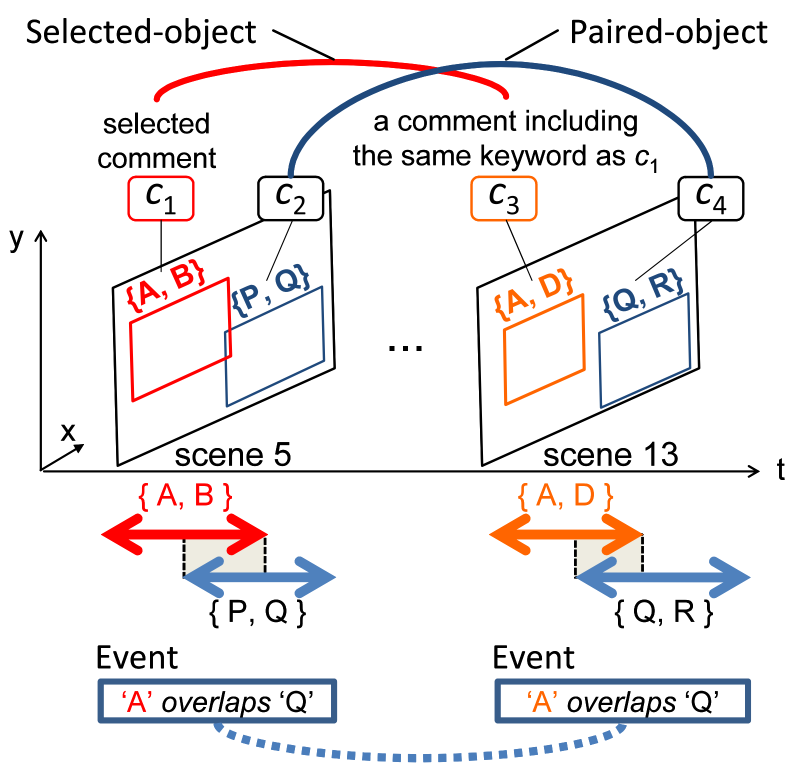


Figure 10: Concept Image of the Determination of Relation between Scenes

3.4.2 Scene Retrieval Application

For a user who views a video clip and is interested in an object in the scene, our proposed system is able to retrieve scenes of comments that refer to the same object by using the scene relation *EQUAL* and *OBJECT*. Even if a comment does not contain the keywords in the selected comment, the system will still retrieve relevant scenes if the comment refers to the same object. Our system can also determine when comments containing some of the same keywords refer to different objects by comparing the comments' pointing regions.

As most video clips are not annotated, we consider it is useful for the user to be able to comment on and view scenes depicting an event of interest and thereby obtain more information about the event. Our system is able to retrieve scenes tagged with comments that refer to the same sequential event by using the scene relations $EQUAL$ and $EVENT$. Comments regarding different objects in the same event are located through a comparison of the $EVENT$ relation. The system retrieves scenes depicting the same event even when comments about the event refer to different objects.

3.5 Evaluation

3.5.1 Prototype System

We developed a prototype system based on our proposed method using Microsoft Visual Studio 2008 C#. This system consists of a video-viewing interface, a comment-posting interface, and lists of relevant comments on objects and events as illustrated in Figure 11. A user can use the comment-posting interface to attach comments with specified pointing regions and temporal durations while viewing a video clip. In the video-viewing interface, the user can view the video clip with user comments that are displayed on the screen at each starting time. As the user selects a comment by clicking the mouse on a displayed comment, the system extracts user comments related to the selected comment based on the keywords that the comment's text is divided into by the morphological analyzer Mecab [64], which is in Slothlib [65], and also based on the comment's pointing region and temporal duration. The system then lists relevant comments and the user can view the scenes where the same object or event appears as the one in the selected scene.

3.5.2 Experimental Evaluation

We used video clips attached about 200 user comments with pointing region and temporal duration as our experimental dataset. In this experiment, we gave several comments to each of the 4 subjects, and then they chose the comments and scenes they thought were relevant for each given comment.

As shown in Table 3, we evaluated our proposed scene extraction method

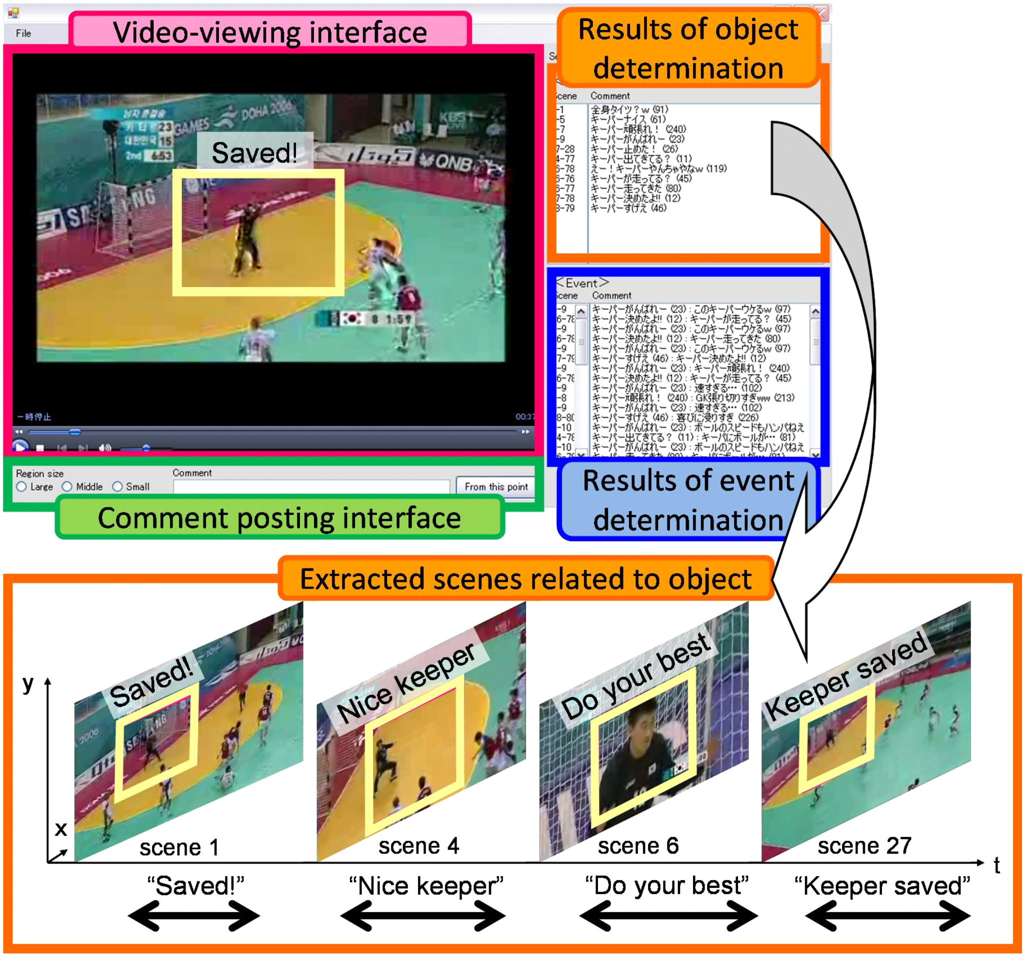


Figure 11: Screenshot of Prototype System

by comparing with a matching keywords method. The accuracy of our proposed object determination based on the relations between pointing regions of user comments was higher than the one of the matching keywords method. Especially, our method can extract comments only on the basis of keywords. In addition, we could show that our system can extract relevant scenes to the users selected one, even though we did not use image processing. On the other hand, the accuracy of the event determination based on the relations between temporal durations of user comments was lower than the one of the matching keywords method. Especially, the value of recall which is an indicator of correctness for the extracted results was exceptionally low. We consider that the result was caused by strict condition of the event identification based on detailed relations of temporal durations. Therefore, we need to integrate corresponding relations,

Table 3: Evaluation of the extracted scenes

	Scenes related to object		Scenes related to event	
	Our system	Matching keywords	Our proposed	Matching keywords
Precision	31%	28%	43%	52%
Recall	65%	58%	26%	35%
F-measure	0.42	0.38	0.32	0.42

though we should consider the granularity of the relations due to types of video clips. However, by utilizing relations of temporal durations of user comments, we confirmed that the system was able to extract scenes similar to scenes the users selected on the basis of identified situations, such as scenes of events give the similar emotions.

In Figure 12, when the user views the video clip of the handball game and selects the comment "Rapid movements" that was written to the goalkeeper about the shot event in scene 12, our system extracts two other scenes as relevant scenes. Both scene 8 and scene 66 are extracted because the scene relation is $OBJECT$, or there are comments that are determined to refer to the same object by the degree of overlap of the pointing regions, "Strange movement" and "Wide movement" in scene 8 and "Strange pose" in scene 66. Scene 8 is extracted also because the scene relation is $EQUAL$, or there is the comment pair "Wide movement" and "GK is so excited," which are considered to refer to the same event as the selected comment and the other comment "Come on, GK!" in scene 12. This comment pair has the keywords 'movement' and 'GK' in common, and the relationship of the temporal durations of the two keyword pairs is the same, $finishes$.

This system extracts scenes of the keeper making an unusual movement during a shot on goal. The images of these scenes are similar. In this way, scene retrieval related to objects can extract scenes with similar images simply by analyzing comment information. We explain an example of scenes that are extracted as the result of the scene retrieval concerning events. In Figure 13, when the user watches the video clip of the soccer game and selects the comment "The keeper's reaction is good" that was written to the goalkeeper about the shot event in scene 21, the system extracts two other scenes as relevant scenes.

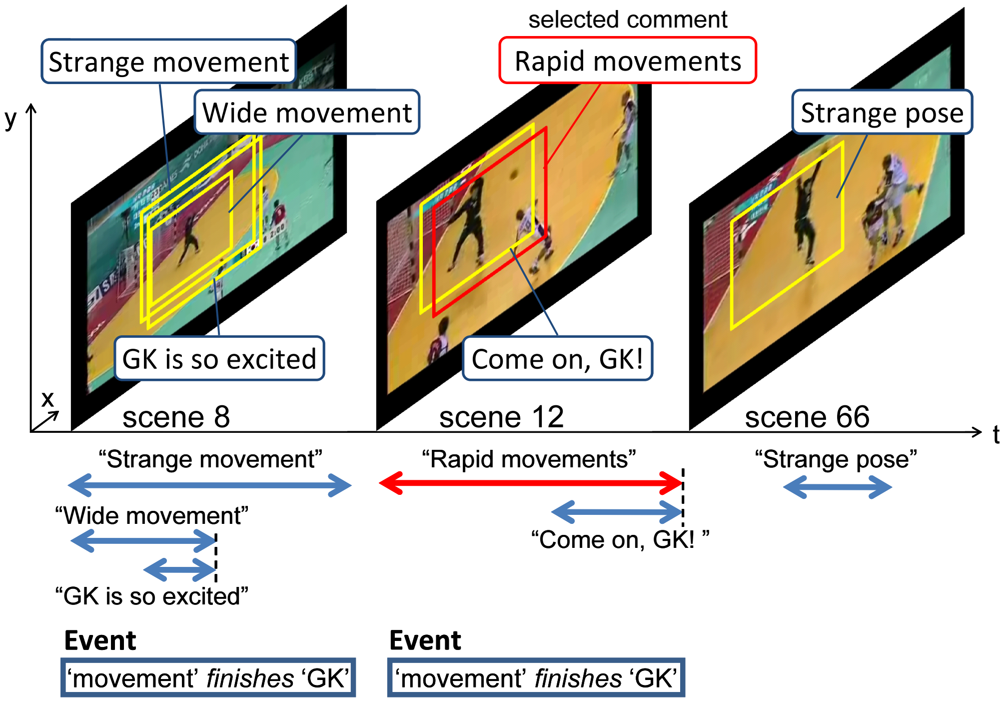


Figure 12: An Example of Scenes Related to Objects

Scene 1 is extracted because the scene relation is $EVENT$, or there are the comment pairs "His reaction is so good" and "Nice save" or "Nice keep" that were linked to the same event as the selected comment and the other comment "Nice shot and keep" in scene 21. These comment pairs have the keywords 'reaction' or 'good' and 'nice' in common, and the relationship of the temporal durations of two keyword pairs is the same, *meets*. Also, scene 16 is extracted because the scene relation is $EVENT$, or there is the comment pair "The keeper isn't moving" and "Stopped" that were linked to the same event as the selected comment and the other comment "Stopped!" in scene 21. Both comments in the pair have the keywords 'keeper' and 'stop', and the relationship of the temporal durations of two keyword pairs is the same, *starts*. Scene retrieval based on events extracts scenes of the shot event to which users have responded with similar reactions. It is especially well suited for videos of soccer, handball, basketball, etc, because sports tend to follow the same event patterns; for example, the shooter kicks the ball, and the keeper tries to stop the ball.

Throughout this experiment, we consider that we have to improve the identification methods for both objects and events to increase precision and recall absolutely. For this, we will try to improve the classification method of the user comments in terms of sentiments such as "enjoyment" and "sadness" or "positive" and "negative" and communities where users belong like channels on YouTube.

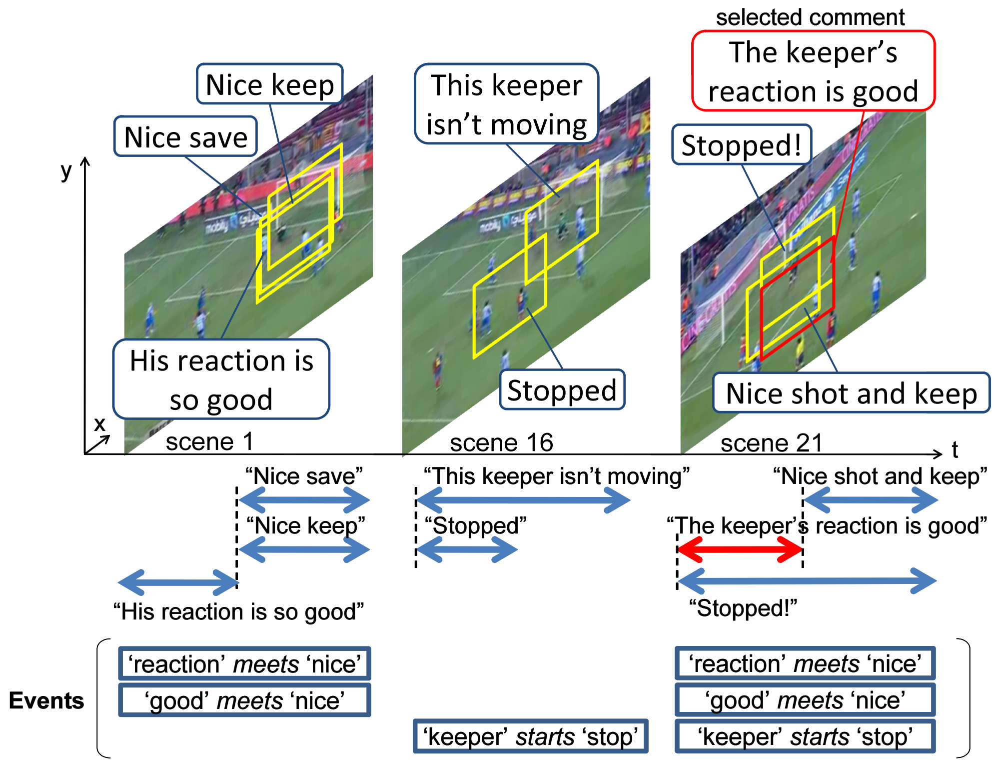


Figure 13: An Example of Extracted Scenes Concerning Events

3.6 Summary

We proposed an method for extracting objects and events based on both the text information and the non-text information of user comments, namely, the pointing region and temporal duration. Then, we presented a method for extracting scenes from video clips based on objects and events as indicated by these comments. We focused on objects and events that are specified by texts and pointing regions or temporal durations of user comments. Our proposed system retrieves scenes according to their relation to one another as defined by our set of relation types: *equal*, *object*, *event* and *no relation*. These are determined by the comments' reference to objects and events in video scenes. In the experiment, to evaluate the extracted scenes, our system could extract relevant scenes with about 200 user comments for each video clip. In the extracted scenes, relevant scenes which were not extracted by the matching keywords method were included. In addition, our search method using comments could extract scenes relevant to a user's interests at lower cost than video analysis. The relevant scenes concerning events were similar to situations in the scenes selected by the user. Consequently, although the accuracy of both determinations are not so good, we could confirm a possibility to exploit non-textual features of user comments for various applications as well as scene extraction.

Chapter 4 Crowd-powered Video Rating: Measuring Relevancy between Tweets and Videos

4.1 Introduction

The recent advances in social networking sites such as Facebook [4] and Twitter [3] encourage crowds to share their updates in almost real time across the open space. At the moment, a new kind of interaction between the TV stations and general audiences increasingly appears stimulating beneficial interactions between both sides. Generally, in the side of TV stations, they want to listen to their audiences' opinions on their contents. Conversely, audiences would like to often participate in the TV program expressing their thoughts or feelings directly to the content providers. Accordingly, in terms of conventional TV viewing surveys, social media must be a valuable source to gather much bigger and wider audiences rating, with less additional costs to selected participants who worked for the conventional TV ratings.

In fact, current TV ratings in the USA and Japan are measured mostly based on Nielsen ratings, which were developed by Nielsen Media Research [66] many years ago. This method measures TV ratings in three different ways: First, "Set Meter," which means an electronic device to monitor what TV programs the selected homes view. The collected viewing logs are transmitted in the night or in real time to the Nielsen center or other media research companies to derive a statistical summary. Next, "People Meter," which is a specially designed remote controller, to recognize the members of a household who watch the TV programs by selecting one of the identification buttons on the remote controller, eventually enabling analysts to survey various demographic groups such as younger vs. older generations. Lastly, "Viewer Diary," the oldest way, is based on audiences' self-recording on paper-based questionnaires about what they have watched individually. The first two methods, which are most often employed methods, need to have the specific devices set up on television sets. Apparently, unlike a few decades ago, we are no more bound to watch TVs at our homes. We can carry the TVs to any outdoor place through smart phones. Additionally, watching TVs is also not limited to the broadcasting time, rather, video recorders or the recently introduced time shift functions in TVs can help us make up for the missed programs consequently. Furthermore, the concept of TV is now extending its realm with many on-line video sites such as YouTube and Japan's NicoNicoDouga. The increasing scope of watching programs on the TV makes it much difficult to assess the ratings through home-centric measurements.

In order to overcome the above problems of conventional TV rating methods, we focus on a new source by crowds. Obviously, among the numerous postings on today's social networking sites, there are many useful crowd lifelogs related to media consuming. In practice, recent TV programs are adopting Twitter as a backward-channel to directly obtain audiences' opinions about on-aired programs. In a sense, this movement can be seen as a form of interactive TV. Based on this conception, we may trace personal media life patterns from these logs and probably rank TV programs or songs that audiences mostly prefer. For example, in the case of YouTube, the most popular video sharing website, their pages have a tweet-writing button to each video viewing page to let people remain or share their viewing experience.

In this paper, we present a novel rating method for targeting several video media such as TV programs and on-line video clips by utilizing the audiences' media lifelogs over representative microblogging site, Twitter, as shown in Figure 14. However, the site was not designed for this specific goal to collect the TV-related Twitter messages, so-called tweets, identifying those that are relevant to TV programs, which are the target of this work. We also need to check the tweets to a particular region of broadcasting in order to filter out other tweets from regions outside our interest. In particular, we will present a semantic linking between tweets relevant to TV programs.

4.2 A Twitter-based Video Rating Platform

In this section, we first describe our motivation to utilize Twitter as a back-channel from audiences to broadcast stations realizing the so-called interactive

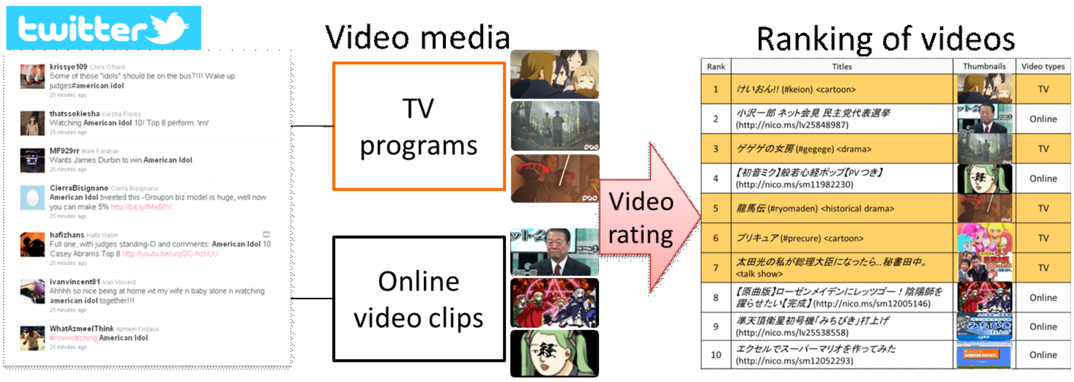


Figure 14: Video Rating Method using Twitter

video media. Then, we introduce a video rating platform on Twitter and highlight the most critical issue of constructing semantic links from tweets to relevant videos.

4.2.1 Looking for Audiences on Twitter

Generally, Twitter is a microblogging site targeting for various real life applications. Thus, in order to use the Twitter platform as a back-channel for video rating, we should find out tweets related videos. In order to look for those TV-related tweets, we may use some specialized hashtags which are popularly used on the site as an index to enable retrieval by other services or users. Generally, Twitter users can simply create a hashtag by prefixing a word with a hash symbol "#hashtag." For instance, various hashtags, such as "#tvasahi," "#2010wc," "#worldcup," "#jfa2010," and "#wcj," "#wc2010" had been used during the 2010 FIFA World Cup. Actually, twtv.jp [67] has already used these tags to collect the messages sent intentionally to the stations. However, the hashtag-based video rating method is not sufficient to aggregate large amounts of public opinions, since hashtags are not always given for all existing TV programs and it needs an effort to intentionally add specific tags relative to TV stations or programs in the current situation where users must write hashtags in a tweet with different devices such as smart phones or PCs during their watching. Furthermore, it is unlikely that all existing broadcast stations have their own Twitter accounts and hashtags. Therefore, it is important to develop a method to capture much more TV-related messages. In fact, this requires a kind of semantic linking between the freely written messages under the length limitation of 140 bytes and the TV programs. We will present the details of our method to link tweets to the corresponding programs, if there are any relevancy between them.

4.2.2 Twitter-based Video Rating Platform

In order to develop a video rating system utilizing crowd lifelogs about several video viewing on Twitter, we propose a tweet-based video rating platform as illustrated in Figure 15. With this platform, we support analysts enabling them to easily investigate crowds' media consumption. Especially, since we are looking into the video viewing logs on Twitter, we need information on on-air programs and on-line video sharing sites to identify what crowds are looking.

Furthermore, to realize monitoring to local video ratings, we are dealing with geo-tagged tweets which have information on when and where a tweet is written. For this, we developed the geographic tweets monitoring system in our previous work [43] intentionally to collect such specific type of tweets effectively. For the simplicity, we will not describe the detail of the system and methodology here. Instead, with a geo-tagged tweet database collected by the system, we will investigate the video-relevant ones and utilize them to populate videos from several video media. Specifically, when we identify the most relevant video with a tweet, we approach two different levels of identification processes. First, as a primitive and essential step, we look up a prepared hashtag list which includes hashtags and the source information. However, as aforementioned, hashtag-based linking from tweets to relevant programs will eventually suffer the lack of relevance enough to measure final video ratings, because unlike on-line TVs or video sharing sites, on-air TVs required a user effort to manually write such hashtags into the writing tweets as shown in Figure 16. In general, people prefer to write just a title of program or a few keywords representing program. In other words, we cannot ignore such freely written texts which are connected to much more hidden video audiences. Therefore, we further have to semantically examine the relevance between tweets and possibly relevant programs from those raw texts. As for on-air broadcasting TV programs, we use an Electronic Program Guide (EPG), which typically provides people with scheduling information for current and upcoming programs.

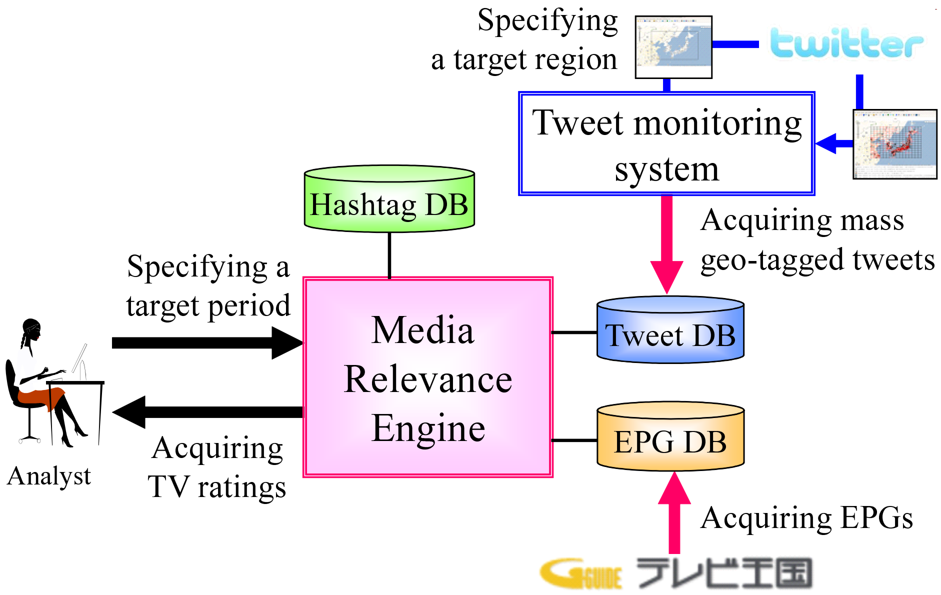


Figure 15: Tweet-based TV Rating Platform

4.3 Semantic Linking from Tweets to Relevant TV Programs

In order to construct semantic linking from user-written tweets to possibly relevant TV programs, we analyze tweets in two different levels of processing. For this, we developed a Media Relevance Engine as depicted in Figure 17. First, for each tweet, it goes the first relevance assessing by checking included hashtags. In this stage, we use a list of hashtags, where each one consists of triple attributes of <"hashtag," "station," "program"> as shown in Table 4 (a). However, all these tuples do not need to be filled, since a hashtag can only refer to a TV station or a program. Additionally, in case of on-line videos, the "station" attribute will specify the site name. In the next stage, "term-based identification" step will identify TV-relevant tweets by examining where each one has some specific strings which can determine a station or a program uniquely. For instance, a partial string of "http://www.youtube.com/..." will give a hint that it is a tweet relative to a YouTube video as described in Table 4 (b).

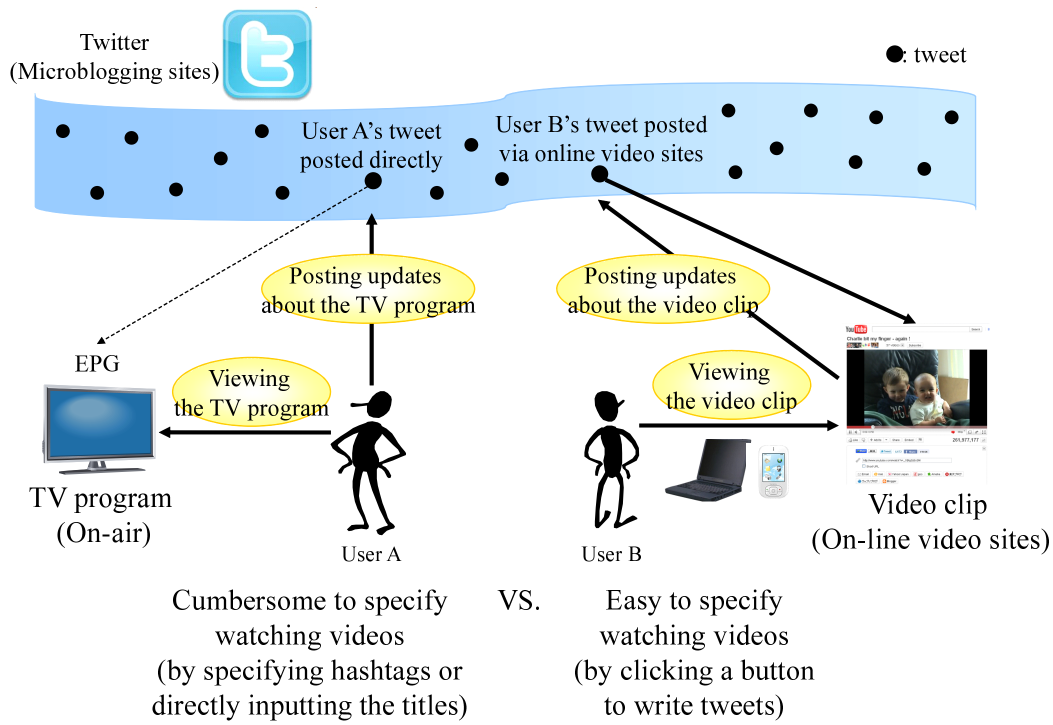


Figure 16: Difference of Crowds' Behavior for Viewing On-air TV Programs and On-line Video Clips

However, after this step, there are still lots of unidentified tweets which are obviously failed to determine any relevance as a TV-relevant tweet, but they can be if we look into the content in detail. For this, we present a Semantic Media Linkage Engine to examine the tweet's media relevance by content analysis as illustrated in Figure 17. In this part, we will find semantic relevance of tweets to on-air programs appeared on the EPG lists. As for EPGs, the items that were broadcasted during the specified period are obtained from the local EPG database, which has also been storing EPG items from TV Kingdom [68] as shown in Table 5. We use these tweets and EPG items as the datasets of the specified period. Then, we need to compute the relevance between tweets and EPGs. Basically, for a tweet relative to a TV program, we have to find the best matching case in the EPGs. Since tweets in this stage, we have to exploit other information on textual, spatial, or location information of tweets.

Actually, the other information is all required to assess the relevance in a comprehensive way. For instance, will be in hashtags identification, as drawn in Figure 18, a user is location in the middle of a city and these are four different broadcast stations around there. But only three stations tv_a, tv_b, and tv_c are accessible from the location of the user. If a tweet written by this user is matched with some program information broadcasted from the surrounding four stations, we can think that the user's message can be to these programs. However, the station tv_d cannot support this assumption, since it is out of the period. Furthermore, in terms of broadcasting time, it is likely that the programs broadcasted in the nearly same time range with the written time would be desirable. Therefore, we need to compute the relevance of tweets to find out relevant on-air programs in the respects of textual, spatial, and temporal relevance as follows.

Table 4: Example of Hashtags and Key-terms

(a) Example of hashtags used for on-air TV stations or program

Hashtag	TV station	TV program
#nhk	nhk	-
#tbs	tbs	-
#tvtokyo	tvtokyo	-
#etv	nhk_edu	-
#keion	-	Keion!
#precure	-	Precure
#gegege	nhk	Wife of gegege
#ryomaden	nhk	Ryomaden

(b) Example of hashtags and key-term lists for on-line video sites

On-line video sites	Hashtags, terms, and URLs for Linkage
nicovideo	#nicovideo
	niconico
	http://www.nicovideo.jp/watch
YouTube	#youtube
	YouTube
	@YouTube
	http://www.youtube.com/watch
DailyMotion	#dailymotion
	dailymotion
	http://www.dailymotion.com/video
Gyao	#GyaO
	@Yahoo_Gyao
	http://gyao.yahoo.co.jp/player
Veoh	#veoh
	veoh
	http://www.veoh.com/browse/videos

Textual Relevance (TeR) In order to find a corresponding EPG item relative to a tweet about a TV program, we first applied a words-based sim-

Table 5: Example of Local EPG Database

region	station	date	time		title	genre
			start	end		
CATV Tokyo area J:COM Tokyo (Suginami)	NHK General Tokyo	Sep. 1, 2010	0:00	0:15	News and weather information	news
CATV Tokyo area J:COM Tokyo (Suginami)	NHK General Tokyo	Sep. 1, 2010	1:05	1:50	Chase! A to Z	documentary
CATV Tokyo area J:COM Tokyo (Suginami)	NHK General Tokyo	Sep. 1, 2010	1:50	2:00	Scoop! Contributed video clips (Tokudane! Toukou DO-ga)	talk show
CATV Tokyo area J:COM Tokyo (Suginami)	NHK General Tokyo	Sep. 1, 2010	2:00	2:45	Try and convince (Tameshite Gatten)	talk show / lifestyle

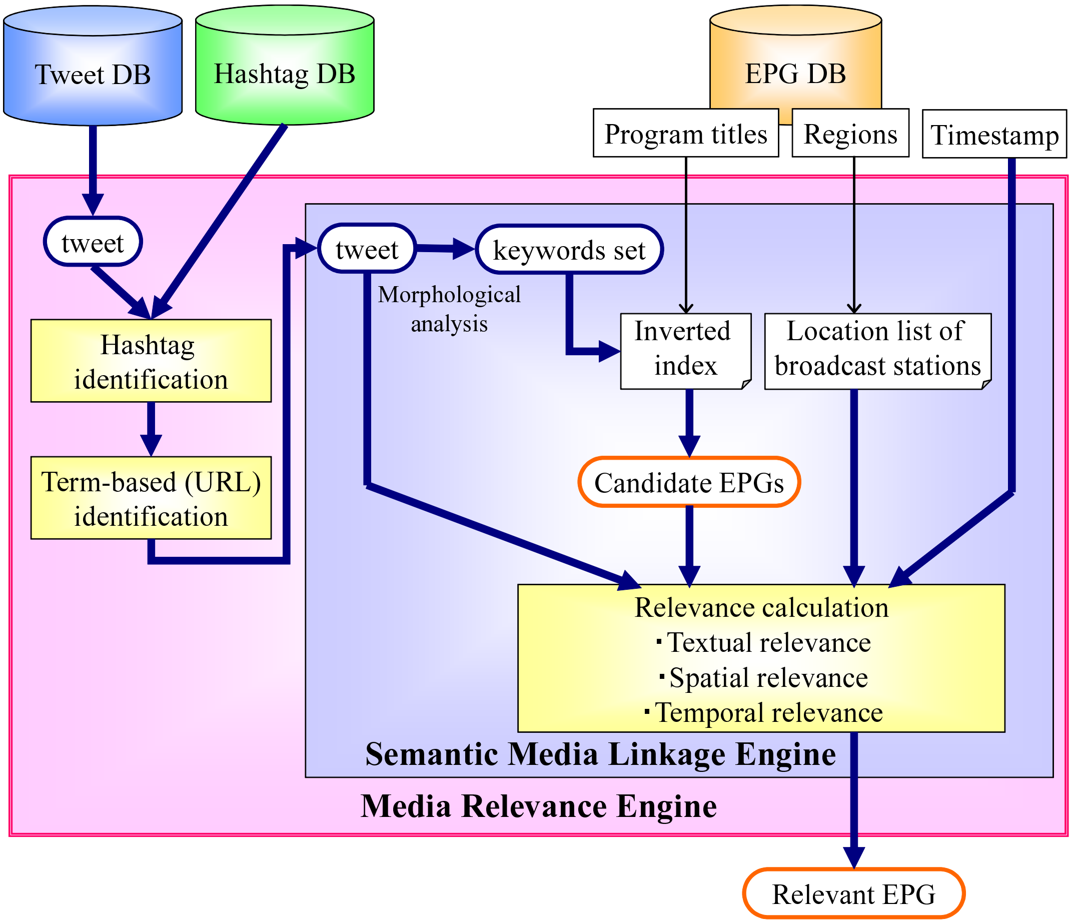


Figure 17: Detecting Relevance between a Tweet and EPG

ilarity computation: both sides are textual message. In the estimation of the correspondence, we compute it with the following formula based on the Jaccard similarity coefficient [69], where tw_i is a tweet, e_j is an EPG item, $e_j.title$ is the title in the EPG item, and mp is a morphological analysis

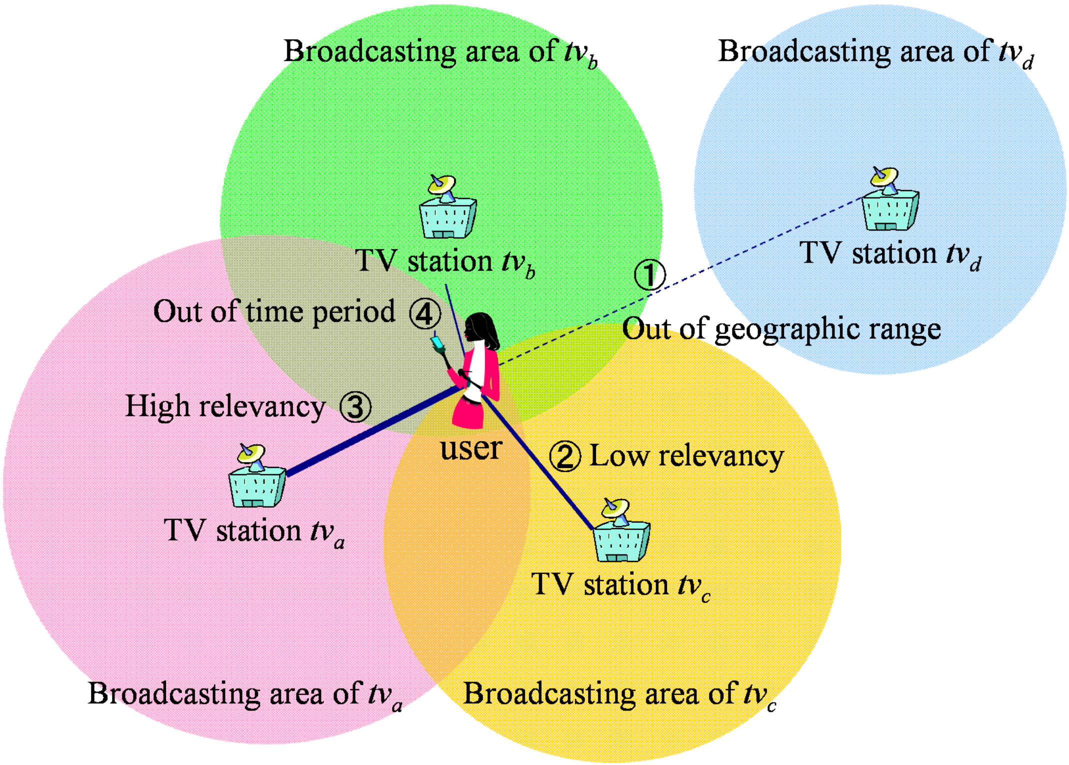


Figure 18: Computation of Semantic Linkage by Textual, Spatial, and Temporal Relevance

function where the output consists of nouns found in the given message. *df* is a function that calculates document frequency. Each tweet should be compared with all the program titles in the local EPG database. For the rapid searching for seemingly relevant EPGs we use an inverted index [70] to reduce the number of calculations required for determining relevance between a tweet and program titles in comparison to directly using the EPG database wherein the total of possible combinations would be enormous. Then, in order to detect relevant EPGs related to titles of TV programs, we applied the formula (10) in the computation. In the formula, with the *df*, we also considered the frequency of keywords of EPGs' titles. For example, keywords that are frequently used in EPGs such as "news," "drama," and "sports" should have less weight since these generic terms would retrieve many unrelated EPGs.

$$TeR(tw_i, e_j) = \frac{|mp(tw_i) \cap mp(e_j.title)|}{|mp(tw_i)| + |mp(e_j.title)| - |mp(tw_i) \cap mp(e_j.title)|} \times \Sigma_k^{|mp(e_j.title)|} \frac{1}{df(k)} \tag{10}$$

Spatial Relevance (SR) According to EPG items in the local EPG database, the same titles of EPGs are often found, because some TV programs can be broadcasted repetitively by multiple stations. In this case, we should identify the station that broadcasted the program at the time of tweet occurrence. The number of TV programs extracted by the inverted index usually corresponds with many different local stations. However, a user can exist at a place in a given moment so that a TV-relevant tweet should be matched to one of the possible local stations. Therefore, we should consider the physical distance between the location where a tweet is posted and that of the station that broadcasted the TV program. Because specific locations of stations are not included in EPG items, we roughly estimate their locations based on "region" attributes of the EPG items using Google Maps API [71]. For this, we use the stations' location list that was generated beforehand. Then, we calculate distances between a location where a tweet was posted and each station, and the station that has the minimum distance is selected.

$$SR(tw_i, e_j) = log(euclid\_dist(tw_i.loc,\ e_j.tv\_station_k.loc) + 1) \tag{11}$$

Temporal Relevance (TR) There is also an important consideration regarding the tweet posting time. Usually, we can think that TV-relevant tweets may be written near the actual on-air time. For instance, audiences may write a lot of tweets during or just after a popular drama. Sometimes, before a very popular sports program such as the World Cup, many tweets may occur far before the actual on-air time. Therefore, as regards the relevance between tweets and TV programs, the time elapsing between them is also an important factor.

$$\mathrm{TR}(\mathrm{tw}_i, e_j) = \begin{cases} 0 & (e_j.start\_time \leq tw_i.time \leq e_j.end\_time) \\ log(e_j.start\_time - tw_i.time + 0.1) & (tw_i.time < e_j.start\_time) \\ log(tw_i.time - e_j.end\_time + 0.1) & (e_j.end\_time < tw_i.time) \end{cases}$$

Final Rating by the Triple Relevance Measures Based on the above criteria, we computed the final relevance using the following formula:

$$relevance\_score(tw_i, e_j) = \frac{TeR(tw_i, e_j)}{SR(tw_i, e_j) \times TR(tw_i, e_j) + 1} \tag{12}$$

After computing the relevance scores, we obtain a list of tweet-EPG mapping and populate TV programs based on the following popularity score. In this formula, $\#tweets$ denotes the number of tweets for a program e_j, while $\#users$ means the distinct number of users. In fact, we consider the biases occurring by aggressive users to write many tweets for a program should be normalized.

4.4 Experiment

4.4.1 Dataset

In order to achieve our purpose to rank TV programs by means of Twitter users, we prepared a dataset for a period between Sept. $1-30$, 2010: 1) tweets that occurred in that period in Japan, and 2) EPGs of all TV stations (except CS satellite broadcast) in Japan. In that period, we collected $6,276,769$ geo-tagged tweets, which were all mapped onto location points on a map. However, it was still burdensome to use this tweet dataset in our preliminary test. For the practical findings of TV-relevant tweets, we empirically chose tweets whose relevance to TV watching was seemingly higher using the prepared hashtag lists (for on-air TV programs and on-line videos) and a set of filtering terms such as "テレビ," "TV," "てれび" Japanese expressions for "television," and "視聴," "番組," "見てる," "見ている" expressions for "watching" or "viewing." By filtering using these terms, we could obtain a reduced tweet dataset (119,575 tweets, about 1.9% of the collected dataset). These potential tweets were written by

33.392 distinct users (on average, 3.58 tweets were made per a user.)

In our experiment, we identified TV-relevant tweets and successively ranked the TV programs. In addition, we also prepared $838,636$ EPGs for the same period. The TV program list we compiled covers 110 geographic regions in Japan with 188 different TV stations. (Here, for nationwide stations such as NHK, which may have many local stations appearing in the EPGs, we dealt with them all as different channels for convenience.) On the list, $24,841$ distinct TV programs were identified (actually, 188 unique TV stations exist, but with the combinations with different regions by region, TV stations we could virtually determine 875 different channels.) Hence, each station has 31.9 programs a day (during on air-time) on average.

4.4.2 Results

In the first place, we extracted $60,318$ tweets by means of hashtag identification. Among them, we analyzed what TV stations and programs were popularly referred to by hashtags. For the situations, we could obtain an expected result of shown in Figure 19 where NHK station (tagged by #nhk) than a half of the total hashtags about stations. Likewise, we could find other major TV stations such as Nippon Network Television Corp. (#ntv), Nippon Network Television Corp. (#tbs), and tv asahi (#tvasahi). In addition, for on-air programs detected by hashtags, we could also obtain the result as shown in Figure 19 (a), where the most popular programs are ranked; Keion! (#keion), Wife of Gegege (#gegege), and life history of Ryoma Sakamoto (#ryomaden). Furthermore, for on-line video sites referred by hasetags and specific URLs, we found an interesting result; the most popular one was NicoNicoDouga and YouTube was ranked in the next. Lastly, we made a comprehensive ranking for on-air programs and on-line videos as shown in Table 6. For on-air programs, we investigated them into two types of "on-air hashtag" identified by hashtags and "on-air". As for on-line video clips, we focused on the very detailed URL's which are usually directing a unique video page corresponding to a program in on-air TVs. In the table, we listed the result in a decreasing order of popularity scores. As a result, 7 of the top 15 popular programs came from on-line videos and the others were related to on-air programs. Especially, 3 of the results were found by the aid of extended searches through our proposed semantic linkage process. Consequently, we could say that the proposed method could show integrated ranking of videos from several video media by finding out hidden audiences who were not using hashtags.

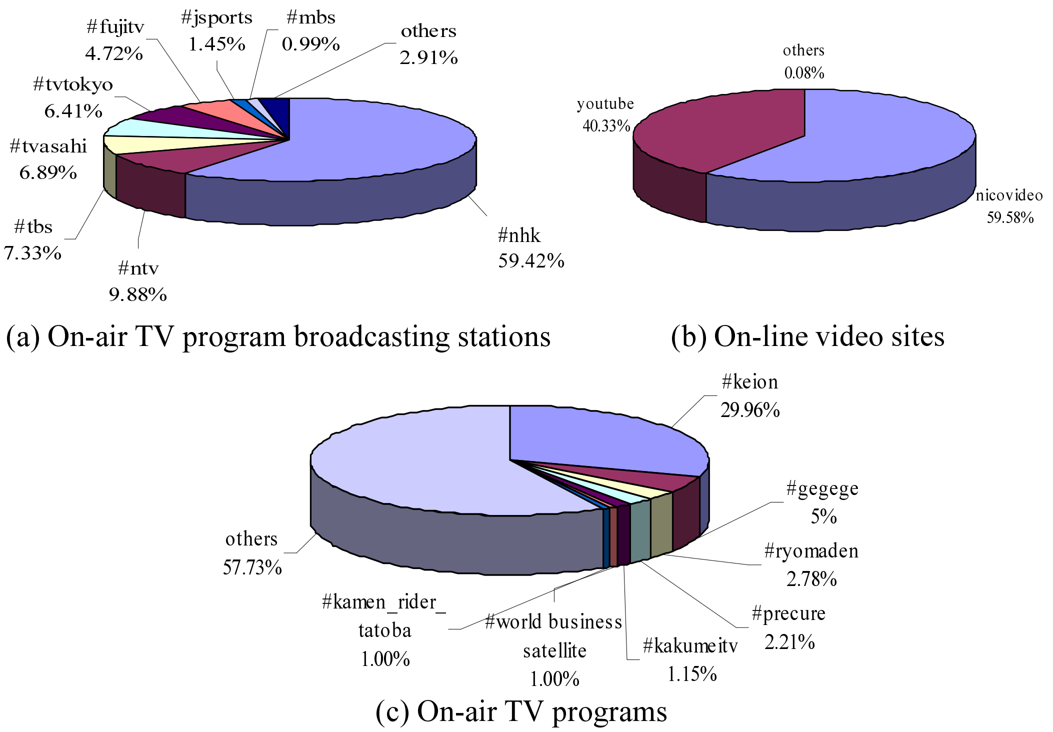


Figure 19: Ranking Result of a) Broadcasting Stations, b) On-line Video Clips, and c) On-Air TV Programs based on Popularity

4.5 Summary

In this paper, we introduced a novel approach to improve a TV rating by borrowing crowd-powered media consuming logs via Twitter. Especially, we also provide a platform looking for audiences of on-air TVs and on-line video clips together. We furthermore presented the very detailed methods and experimental results based on a real dataset of tweets and electronic program guide. In the experiment, we used 119,575 geo-tagged tweets and 838,636 EPG items (24,841 distinct TV programs) or on-line video clips for measuring relevance between

Table 6: Ranking Results of Twitter-based Video Rating for On-Air TV Programs and On-line Videos

programs	type	#tweets	#users	popularity
#keion (anime)	on-air hashtag	978	141	4409.49
http://nico.ms/lv25848987	on-line	147	70	848.70
#gegege (drama)	on-air hashtag	138	59	693.09
http://nico.ms/sm11982230	on-line	57	56	422.79
#ryomaden (drama)	on-air hashtag	76	47	409.74
precure (Anime)	on-air hashtag	145	27	325.12
If I become a prime minister (talk show)	on-air	45	44	295.16
Keion! "Examination" (Anime)	on-air	45	43	288.45
http://nico.ms/sm12005146	on-line	39	37	231.06
http://nico.ms/lv25538558	on-line	61	26	203.07
http://nico.ms/sm12052293	on-line	33	32	183.83
Leading actor —setting beaty salon— (comedy)	on-air	32	30	169.71
#kakumeitv	on-air hashtag	59	22	168.99
http://nico.ms/sm12098837	on-line	31	28	155.90
http://nico.ms/sm12017331	on-line	30	28	153.36

a tweet and a TV program or on-line video clip in terms of textual, spatial, and temporal similarities. Consequently, we could show novel ranking by taking into several video media consideration. In the future work, we will explore much deeper crowds' media lifestyles and their opinions to media contents to activate fruitful interactions between TV content providers and audiences by opinion mining and sentiment analysis for tweets.

Chapter 5 Urban Area Characterization based on Crowd Behavioral Lifelogs over Twitter

5.1 Introduction

Lifelogging increasingly becomes one of our common and daily habit undisputedly exemplified by today's social network sites. In fact, from the early work by Steve Mann's laborious life logging with wearable computing systems [72] to Gordon Bell's MyLifeBits [73] for digitizing every moment of individuals, storing and recalling our lifetime memory have been intensively studied well. However, the recent advances of social networks encourage us to write our lifelogs much easily and share them instantly with any other people around the world. Accordingly, individual lifelogs are not bound only to personal memory. The shared memories of enormous crowds over the open space are extending the individual lifelogs to community experience logs, which can vividly reflect many important social and physical real-world events or phenomena. Specifically, on behalf of the rapid distribution of smartphones and the location-based microblogging sites such as Twitter [3] and Foursquare [5], we can now share our daily activities as well as our minds instantly from any place clarifying where we are located in the world. In particular, this kind of global trend will be delivering lots of novel applications that can benefit from exploiting the shared crowd lifelogs in terms of the huge volume of geo-tagged data and their heterogeneity of contents about almost every kind of crowd activities in the real world.

In this work, motivated by the fact that crowd's lifelogs over the social networks can include real-world location information with the shared messages, we attempt to analyze urban characteristics from the crowd-sourced data over Twitter. Indeed, it is a critical issue to characterize urban space to support various real-life decision makings in the space. For instance, when we have to look for a house to move in an unfamiliar city, it would require a bothering effort to quickly grasp the living environment while drawing the image of the city in mind such as "Where are the most popular places for living with good educational environments?" or "Where is the downtown area attracting together many people on weekends?" Furthermore, in many practical geo-business or geo-politics applications, we frequently have to examine overall features of local areas as quickly as possible. These kinds of questions we would often face will require lots of efforts to obtain, since the possible answers need to investigate many updating sources about ever-changing information of urban areas. In other words, surveying such characteristics of urban space usually takes huge costs and time involving off-line on-the-spot observation or gathering information by the questionnaire method. Hence, quick and massive scale investigation could not be successfully done due to the extensibility problem such as monitoring massive tourists congregating in a city [74]. Consequently, we are limited in characterizing urban areas if we only depend on the lazy static survey results. Otherwise, we only had to depend on the general conception of urban areas previously formed by the mass media or self-experience or rough statistics from national census data.

However, compared to conventional urban characterization research, location-based social networks definitely have many significant advantages and benefits; first, there are enormous populations around the world who are voluntarily publishing their daily activities, particularly, acting in urban space. Second, the crowd-sourced data have various information from trivial travel experience to crowd's seasonal movement trajectories. These heterogeneous types of data shared over the social networks can help us conduct various quantitative and qualitative research studies and realize practical systems. Therefore, the unprecedented scale of crowd lifelogs in urban space will promise many challenging and beneficial issues in the near future.

In general, crowd behavior in urban area is a critical factor to understand urban space. We can easily imagine the strong relationship between characteristics of the real urban spaces and the activities of citizen. For instance, in terms of space syntax techniques by Jiang et al. [75] which attempted to describe urban spaces by crowd behavior, people expectedly become more active in morning and evening hours intensively in residential districts in the sense of integration. Particularly, in a city which we have usually developed for our convenience, our activities such as commuting, working, shopping, educating, etc. will characterize urban areas clarifying how we behave and live in the real space. Therefore, with crowd's daily lifelogs over social networks, we can conduct urban characterization by observing crowd behavior and finding significant behavioral patterns depicting how crowds are using our urban space.

For this purpose, we present a model to explain and study the current situation relevant to social networking sites as illustrated in Figure 20. Here, we intended to construct much more flexible and elastic model, which can comprehensively reflect on the heterogeneity of the participating entities to explain the current situation relevant to social networking sites; that is, the real space (including real-world phenomena and social events as well as geographic physical environment), crowds (generally, people and their various capabilities of sensing, acting, thinking, emotionally feeling, etc.), and the virtual space (while this work limitedly focused on the Twitter as a representative location-based social network).

In the conventional social network studies, graph-based models [76, 77] are often adopted to focus on the relationship of users on the social networks. However, we think that each entity would require its own modeling and operations (e.g. in case of the real space, various spatial and neighboring concepts are often useful, but harder to consider such thing only in a graph). Especially, we would like to focus preliminarily on the influences from the real space to crowds and their reflecting activities on the cyber space by publishing information. In this respect, we presented a simplified model enough to represent the influential relationship among the entities. Furthermore, in order to examine the dynamic nature of the crowd behavior observable from social network sites. We primarily focus on extracting urban characteristics in terms of crowd behavior in the real world that can be largely available by exploiting geo-tagged tweets from Twitter. Specifically, we compute a crowd behavior feature focusing on temporal changes of periodic occurrence of geo-tagged tweets for a geographic region. We also examine significant crowd behavioral features for reasoning urban characteristics. For this, we experimentally extract geographic crowd behavioral patterns from a large number of actually gathered geo-tagged tweets in Japan and report a comparison with real socio-geographic features of each region as an evaluation.

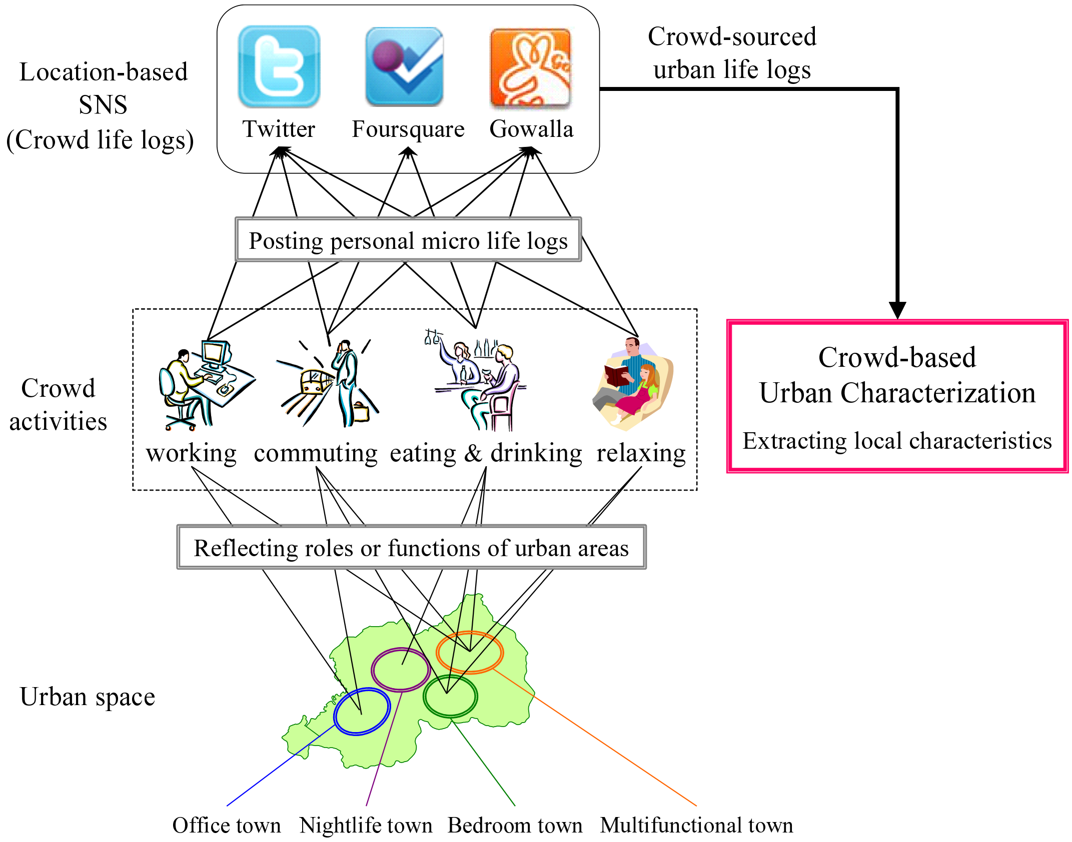


Figure 20: Research Model to Characterize Urban Area using Crowd Behavior Extracted from Location-based Social Networking Sites

5.2 Characterizing Urban Areas with Crowd Lifelogs over Social Networks

5.2.1 How Crowd Lifelogs can Reveal Urban Characterization

Generally, in urban space, there are a lot of facilities for living and working such as housings, transportation, offices, schools, parks and shopping centers. In such complicated space, we can easily observe some daily routines of residents such as commuting, working, eating and drinking, and relaxing at home by exploiting crowd-sourced lifelogs over social networking sites as illustrated in Figure 20. In addition, these crowd activities let us know roles or functions of the urban space;

an urban area which observed crowds who are commuting and working would be conjectured as an office town. On the other hand, if crowds commuting, eating and drinking, and relaxing at home are monitored in the urban area, we might regard there as a bedroom town. Likewise, we are able to capture the image of urban space by means of crowd behavior.

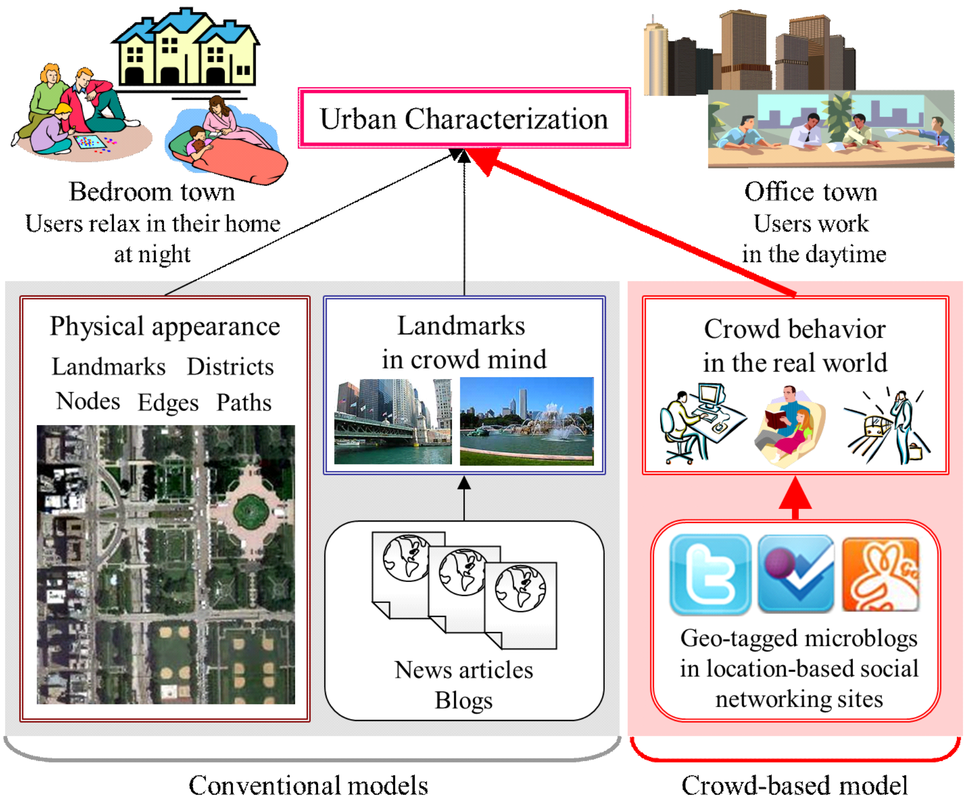


Figure 21: Approaches of Urban Characterization

In fact, geographical characteristics have been studied so far from various perspectives by physical geographic shape or diverse objects such as streets or landmarks, or by cultural and structural aspects such as residential, commercial and industrial districts. These two different views have been well studied in many research fields. Kevin A. Lynch's seminal contribution in his book titled "The Image of the City" [78] defined five fundamental elements of a city; paths, edges, districts, nodes, and landmarks. Based on these elements, Lynch thought that we could characterize our living space within the appearance of a city to imagine ourselves living and working there as shown on the left side in Figure 21. In another remarkable work describing a way to extract geographic characteristics, Tezuka et al. [79, 80] extracted geographic objects and their roles frequently mentioned on the Web contents which can be regarded as a mirror of the crowds' minds to the real world as shown in the middle of Figure 21.

These two different types of urban characteristics may be useful to derive the image of the city. However, in a sense, they are focusing only on static elements of the city and did not take into account the most dynamic and important element of the city; that is, "crowds" living there would be a critical factor for observing urban characteristics. For instance, in the case of the recent terrible disaster, the earthquake and tsunami in Japan on Mar. 11th, 2011, although lots of static characteristics in some part of the devastated Fukushima prefecture area remain unchanged, but the image of the city based on crowds living there was totally changed because the sequel nuclear accidents caused by the disasters prevented people from accessing the towns for a long time.

In this paper, we distinguishably attempt to derive a new kind of social geographic characteristics based on crowds, especially in terms of their behavior, by utilizing the location-based social network sites as shown on the right side of Figure 21. In fact, geo-tagged microblogs over social networking sites can be easily collected since these are shared in the open spaces. We can utilize the free data to extract the image of cities based on crowd activity in the real world.

5.2.2 Modeling Twitter as a Crowd Lifelog Source

Sharing personal lifelogs over social networks such as Facebook [4] and Twitter [3] is a common phenomenon world-wide spreading deeply into our daily lives. Behind the scene of those buzzing trends, we can find some crucial clues to understand why the trivial logs can be very useful to monitor the overall crowd lifelogs. First of all, we can observe three fundamental elements of crowd activity monitoring, that is, user, location and time. These spatio-temporal logs by enormous crowds are possibly appearing from any place where users can write their lifelogs with the help of automatic location sensing functionality of recent smartphones. The implication of this kind of crowd behavior means various facts from simply a possibility of some geo-social event occurrences in a region to a trajectory of a tourist's travel by examining a history of time-variant footprints.

From personal to society level, we can even extract much broader and invisible crowd activity patterns. Of course, the surficial existence of crowds and their moving histories would be the foremost important features we can derive from the massive number of crowd lifelogs. In addition, if we can find much detailed about crowd actual behavior, crowd lifelogs over the social networking sites can be used for various investigation of social trend or pseudo census to people in a city or a nation. We may ask crowd opinion on a variety of social topics simply by referring to crowd messages focusing on some specific words such as the name of political parties or topical keywords. As for the extended crowd activities, we can utilize the written texts by analyzing what kinds of contents are actually written inside. In case of Twitter, though the message field is only bound in term of the length up to the 140 characters, it can have various types of contexts; 1) a hyperlink to external media such as some photo links through "http://twitpic.com" would represent that the correspondent user is probably taking a picture at a place, 2) #hashtag is a promise often used to mean the written textual message is related to a certain topic indexed by the tag, and 3) user networks are extractable, if there are patterns such as @user_id or RT (re-tweet) terms; the first pattern is used for sending a message to specified user, while the second pattern is to represent that the current written message is sourced from other user. Therefore, combining these features with the location and time information can give much more clear picture of what kinds of crowd activities are happening in a geographic area of interest. However, for the direct and straightforward approach, we attempt to model crowd behavior based on three basic metadata of geo-tagged lifelogs over most location-based social networks; time stamp, location information, and user ID, without analyzing textual messages.

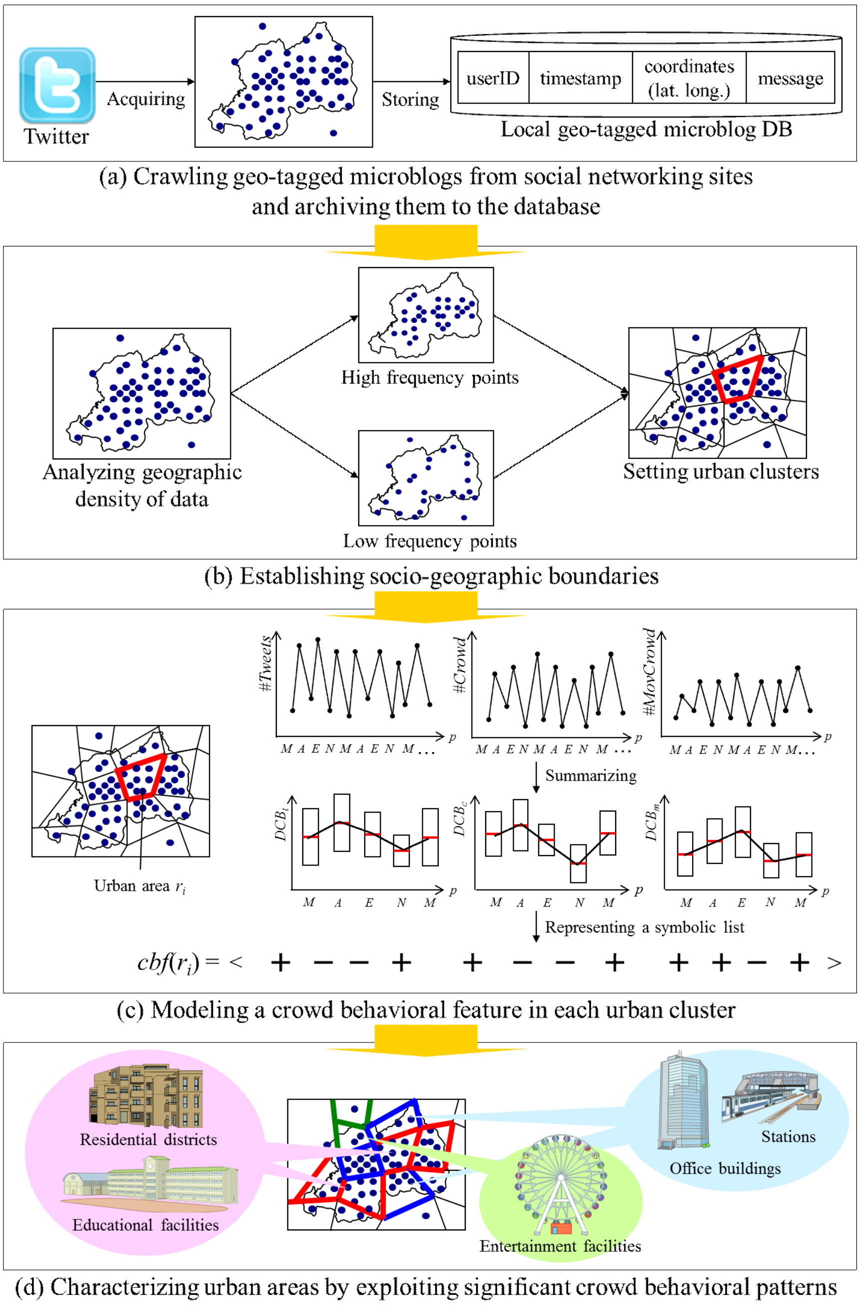
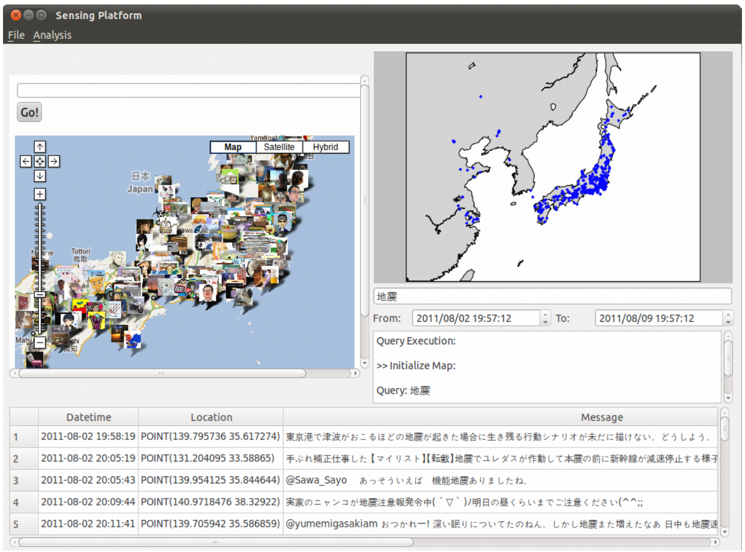


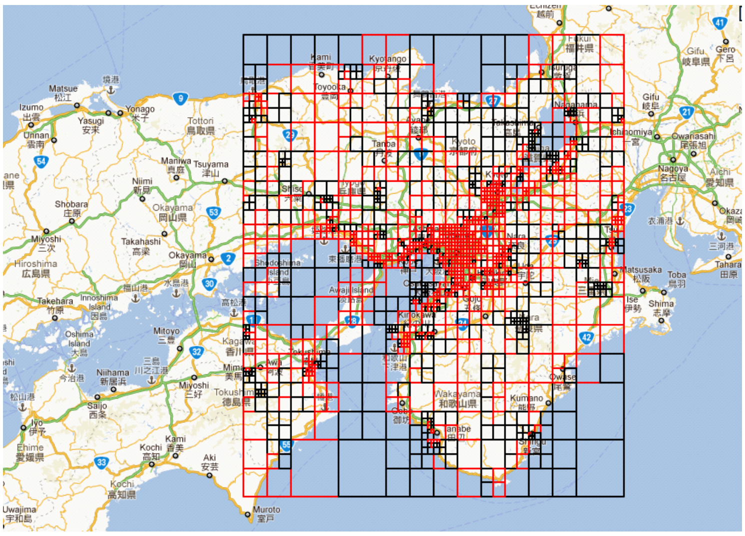
Figure 22: Overview of Crowd-based Urban Characterization

5.3 Extracting Urban Area Characteristics Based on Crowd Behavior over Twitter

In order to achieve our goal of crowd-sourced urban characterization, we describe the further details of methodology about how we can collect crowd's geo-tagged micro lifelogs from Twitter as shown in Figure 22 (a), how we can partition a region by setting socio-geographic boundaries as shown in Figure 22 (b), how we can model crowd behavioral features based on crowd's lifelogs as shown in Figure 22 (c), and how we can extract characteristics of urban areas grouped by crowd behavioral patterns as shown in Figure 22 (d).



(a) Geographic microblog monitoring system



(b) Quad-tree based geographic distribution of tweets around Osaka, Japan

Figure 23: Geo-tagged Lifelogs Acquired by the Geographic Microblog Monitoring System

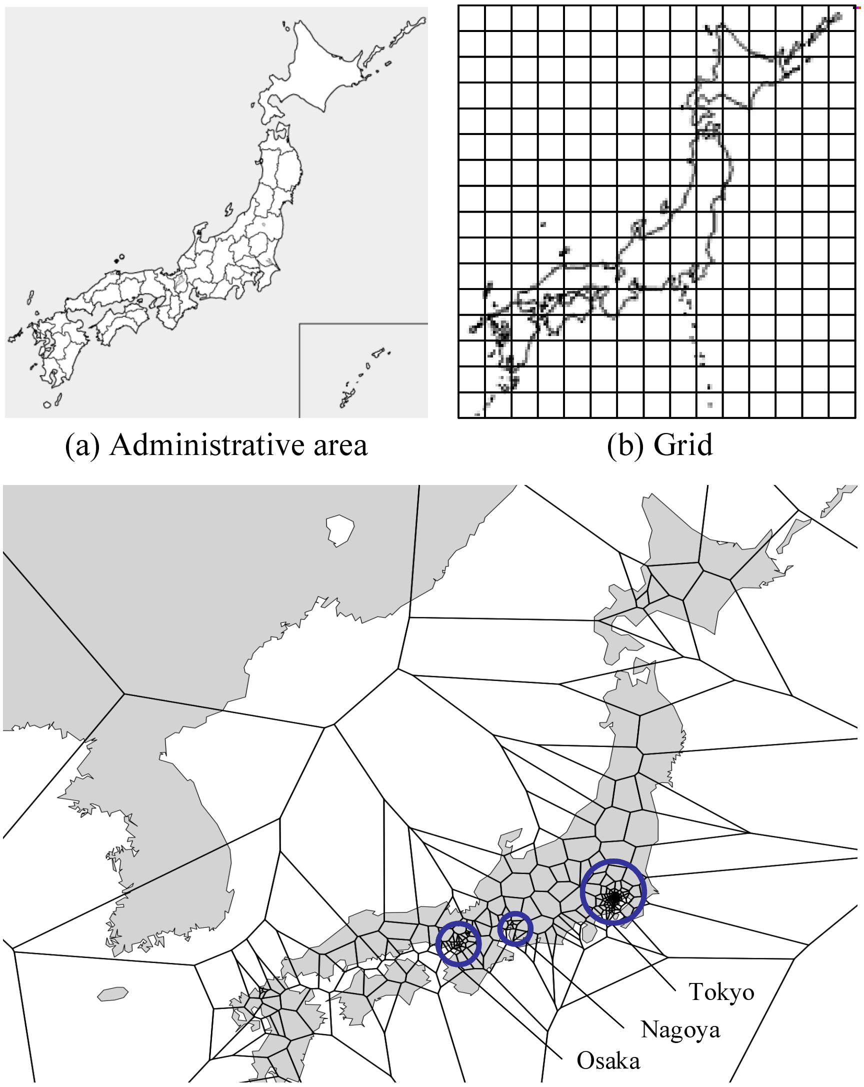
5.3.1 Location-based Social Network as a Source for Crowd Activity Monitoring

First of all, we need to gather geo-tagged tweets from Twitter to observe crowd activities in the real world as depicted in Figure 22 (a). However, it takes a considerable amount of efforts to acquire a significant number of geo-tagged tweets because of certain practical limitations: In fact, Twitter presents open API [81] which solely supports the simplest near-by search based on a specified center location and a radius. Furthermore, each query can only obtain up to $1,500$ tweets for past one week. Therefore, in order to overcome these restrictions and perform periodic monitoring of any size of user-specified regions, we developed a geographic microblog monitoring system [42, 43] which can monitor crowd behavior for a specific region of any size depending on the density of massive geographic microblogs as shown in Figure 23 (a). Figure 23 (b) shows the quad-tree based geographic distribution of geo-tagged tweets from crowds in a part of area surrounding Osaka prefecture in Japan. The location information of geo-tagged tweets can be received either in a raw text form or in very precise location coordinates. Hence, in the case of the former textual style, we needed to perform geo-coding to identify the exact coordinates by translating place names into the corresponding exact locations. We were able to solve the problem easily by using another mash-up service with Google Map's geo-coding service [82]. We directly transferred the place names to this conversion service and received the precise coordinates. Subsequently, we could accurately determine when and where each tweet was written.

5.3.2 Socio-geographic Boundaries of Crowd Activities

Next, in order to monitoring crowd behavior in urban areas for extracting urban characteristics, we need to define urban areas by partitioning a given region into sub-areas.

As for how to partition the region, there are several different space-partitioning; for instance, administrative area, grid, and cluster as shown in Figure 24. Hence, we should select the most appropriate method for our goal; characterizing urban areas based on crowd behavior there. Administrative areas are formed by splitting a target region into prefectures and municipalities based on administrative



(a) Administrative area (b) Grid

(c) Cluster

Figure 24: Space Partitioning for to Monitoring Crowd Behavior

boundaries; limits or borders of a geographical area under the jurisdiction of some governmental or managerial entity as shown in Figure 24 (a). Therefore, it is difficult to figure out if crowd behavior areas are relevant or almost dependent on administrative areas, because crowd behavior often easily cross over the administrative boundaries. Accordingly, we consider that this method might be inappropriate for examining crowd behavior. Next, as for the grid, it is difficult to decide the adequate cell size because a grid is formed by a lot of equal-sized cells as shown in Figure 24 (b). In addition, it would consume considerable unnecessary costs for observing crowd behavior since it does not consider non-uniform distribution of crowds. On the other hand, in case of the clustering, it can reflect the geographical distribution of crowds based on location information of geo-tagged tweets and deal with heterogeneous regions differently. As a result of this approach, small urban areas of densely populated areas appeared around major cities such as Tokyo, Nagoya, and Osaka in Japan as shown in Figure 24 (c). In contrast, large urban areas of sparsely populated areas are spread over the other suburban areas or surrounding sea. Thus, we can effectively establish the appropriate socio-geographic boundaries for the target region and partition into urban areas by referring to the actual geographic crowd behavior.

On the basis of these reasons, in this work, we selected the cluster-based space-partitioning method which can take into consideration geographic distribution of crowds. Especially, in our experiment described in Chapter 5, since we deal with millions of locational data of crowds obtained from Twitter as shown in Figure 25 (a), it requires enormous computational effort. Therefore, we have to reduce the data size in much smaller and computable size without lack of essential quality of the original data. For this, we adopted the NNClean algorithm [83, 84] to split the data into two groups of high-frequency and low-frequency parts as shown in Figure 25 (b) and (c). In many cases, the algorithm is used for distinguishing noise from a given data; low-frequency part. However, in the case of crowd-sourced data over social networks, high-frequency points are naturally observed around high-populated areas. Therefore, using only high-frequency points works out to ignore the suburban areas. In order to solve this problem, we also utilize the low-frequency points. Consequently, we generate clusters from these two different sources respectively by applying EM algorithm [85] as shown in Figure 25 (d) and (e). After that, we depicted a Voronoi diagram [86, 87] using the center points (latitude, longitude) of all clusters and defined the formed polygonal regions as urban clusters as shown in Figure 25 (f).

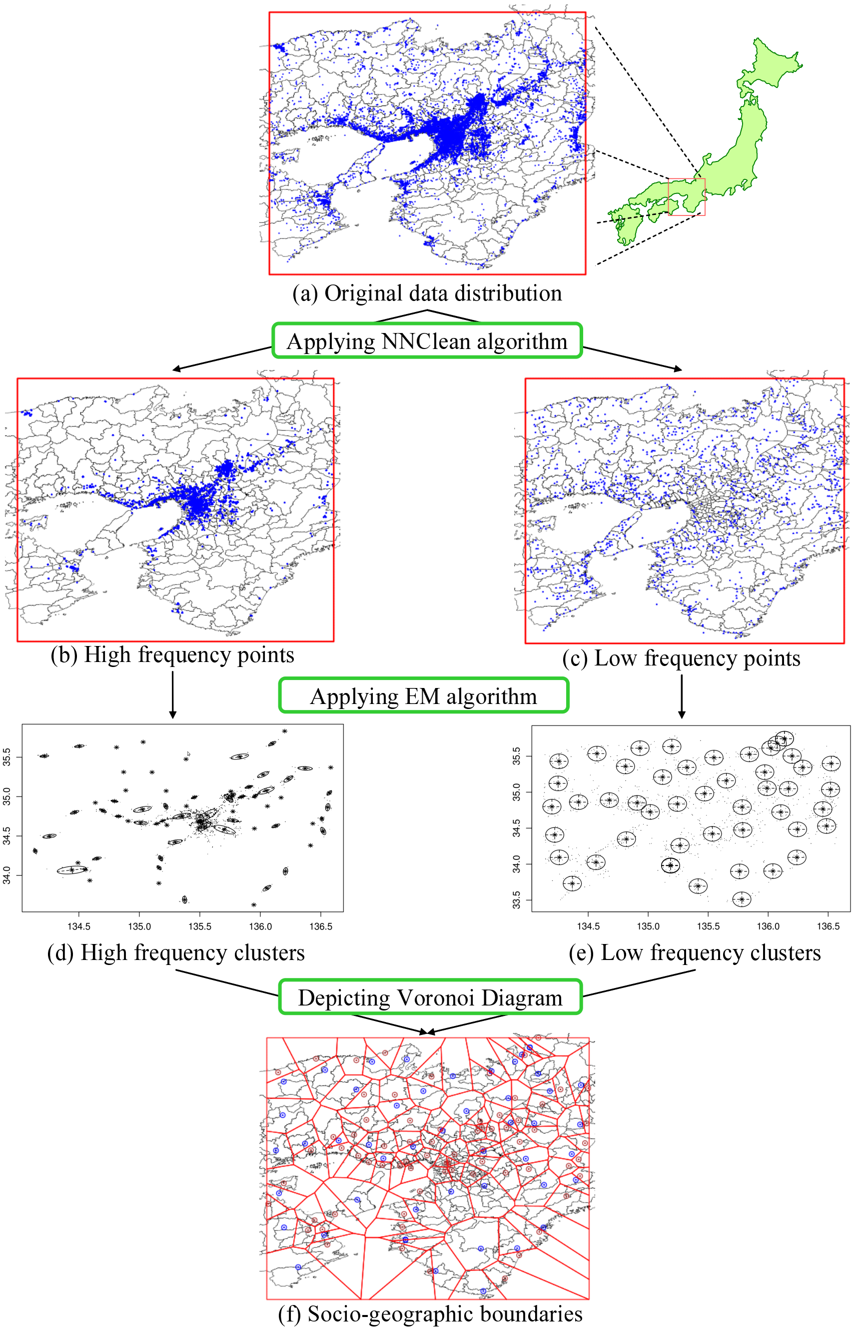


Figure 25: Process of Constructing Socio-geographic Boundaries based on EM Algorithm and Voronoi Diagram

5.3.3 Extracting Crowd Behavioral Features

In order to grasp crowd lifestyles in urban areas, we monitor crowd behavior through their lifelogs over social networks. Specifically, we compute crowd behavioral feature based on parameters computable using primitive metadata of geo-tagged tweets; user ID, timestamp, and location information. In addition, since crowd behavior would vary depending on certain time slots of a day, we should periodically monitor crowd behavior by splitting a day into certain time period p_j. In this paper, we empirically set 6-hour by splitting a day into four equal time slots: Morning (M, $06:00-12:00$), Afternoon (A, $12:00-18:00$), Evening (E, $18:00-24:00$), and Night (N, $24:00-06:00$), and model crowd behavior in terms of three parameters defined as follows:

$\#Tweets|_{r_i,p_j}$: The total number of tweets occurring inside an urban area r_i in a time period p_j.

$\#Crowd|_{r_i,p_j}$: The number of distinct users observable in an urban area r_i during a specific period of time p_j. Naturally, the in-equality $\#Crowd \leq \#Tweets$ is valid since any individual user can write one or more tweets.

$\#MovCrowd|_{r_i,p_j}$: The number of mobile users in an urban area r_i in a time period p_j. To be more precise, this is the number of users posting two or more tweets at different locations in the area. Obviously, the in-equality $\#MovCrowd \leq \#Crowd$ is valid. This parameter is the most dynamic one which can explicitly reflect the temporal usage of real space by crowds.

Crowd behavior feature for an urban area r_i is represented based on these three features; $\#Tweets$, $\#Crowd$, $\#MovCrowd$. Although the scale of crowd behavioral feature would be different depending on crowd in each urban cluster; while periodic occurrence of geo-tagged tweets or the absolute number of crowds are different in each cluster, two or more regions can be similar in terms of an increasing and decreasing tendency, for example, there can be increasingly crowded places such as a large railway station in the morning and the evening. Therefore, in order to explore such significant patterns, we represent crowd behavioral features based on relative temporal changes of the degrees of crowd behavior as shown in Figure 26 (b), where DCB_x refers to a sequence of temporal change between the time slots about the parameter $x \in \{\#Tweets, \#Crowd, \#MovCrowd\}$. Specifically, we compute the differences between the degrees of crowd behavior at two consecutive time slots; $t1$, $t2$, $t3$, and $t4$ from $DCB_t(r_i)$, $c1$, $c2$, $c3$, and $c4$ from $DCB_c(r_i)$, and $m1$, $m2$,

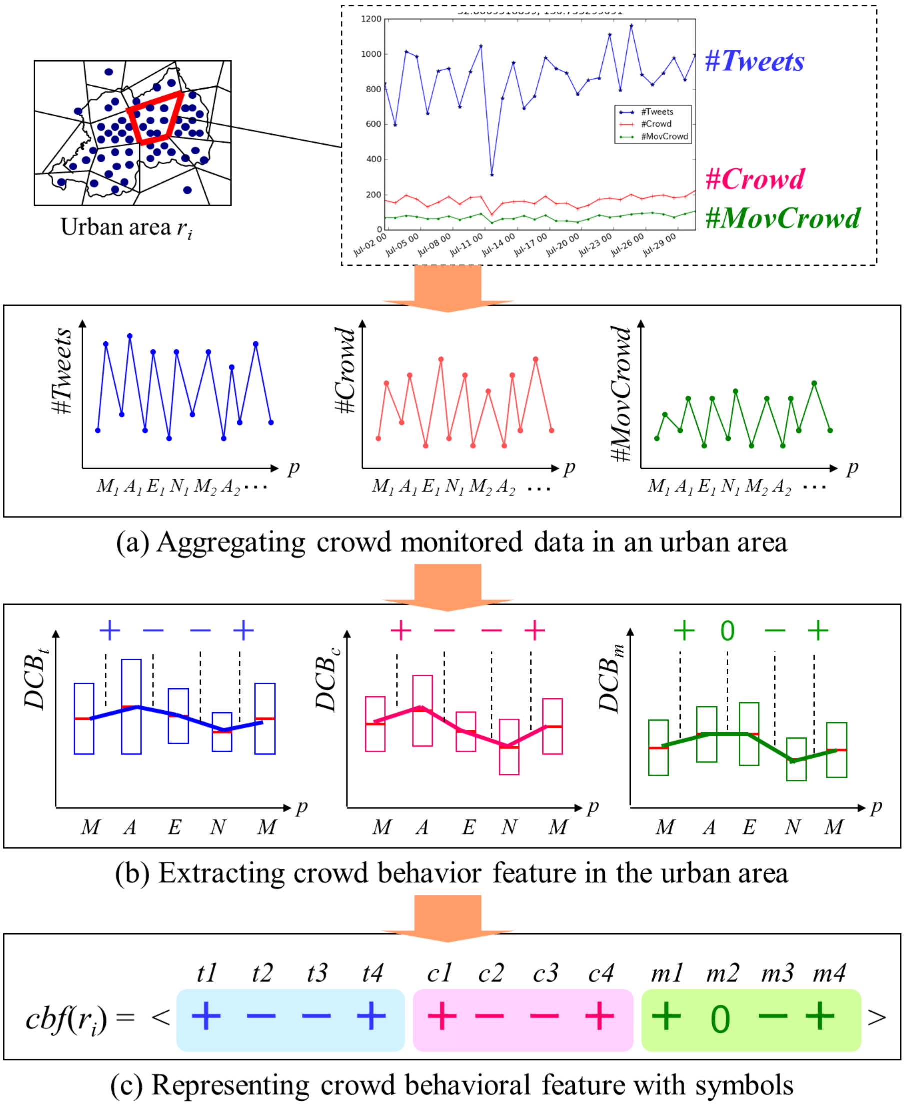


Figure 26: Process of Extracting Crowd Behavioral Features

$m3$, and $m4$ from $DCB_m(r_i)$ in Figure 26 (c). Here, the four suffixes, 1, 2, 3, and 4, represent the differences from morning to afternoon (M–A), from afternoon to evening (A–E), from evening to night (E–N), and from night to morning (N–M). For the simplification, in this paper, we kept crowd behavioral features by only looking up the differences by transforming the representation to a symbol list such as (+, −, 0, +), which respectively means the increase from M to A, the decrease from A to E, no change from E to N, and the increase from N to M. Indeed, the temporal changes can signify the dynamicity of crowd behavior so that we can obtain many possible combinations such as (+, −, −, +), (+, 0, −, +), etc. Finally, based on these symbolic patterns, we define crowd behavior feature, $cbf(r_i) = < DCB_t(r_i), DCB_c(r_i), DCB_m(r_i) >$ as illustrated in Figure 26 (c). This representation will be used as a summary of crowd behavior in a region r_i.

5.3.4 Exploiting Crowd Behavioral Patterns

In the above, we described the method to extract crowd behavioral features in urban areas based on crowd-sourced data. Then, we need to find out significant crowd behavioral patterns monitored in multiple clusters by mining crowd behavioral features. Since each crowd behavioral feature consists of a series of the temporal changes of the degrees of crowd behavior cbf represented by 12-dimensional symbols as depicted in Figure 26 (c), there will be numerous combinations up to the maximum 3^{12}, hence the computational cost to examine the common full-size or partial characteristic patterns would be unbearable. In order to simplify the problem, we adopted a frequent itemset mining algorithm [88] for statistically looking for common frequent patterns from the item sets having a huge size of combinations. Ultimately, we can extract the common characteristic crowd behavioral patterns which can characterize many urban areas in common. In fact, each feature is described by a unique representation, such as $cbf(r_i) = < t1+, t2-, t3-, t4+, c1+, c2-, c3-, c4+, m1+, m20, m3-, m4+ >$ which comes from the original form $cbf(r_i) = < +, -, -, +, +, -, -, +, +, 0, -, + >$, and $t1+$ is a symbol consisting of three factors (‘t’, ‘1’, ‘+’); here, we explain the extended symbols. ‘t’ means that it is a parameter of $\#Tweets$. Next, ‘1’ is an index to distinguish from the others. Finally, ‘+’ is a symbolic representation of the relative change of degrees of crowd behavior during two periods. On behalf of this translation, each symbol can be seen as a unique item, which can be utilized in the frequent itemset mining. Thus, based on this method, we can extract the common partial patterns effectively.

5.4 Experiment

In this section, we describe our experiment to extract significant patterns of crowd behavior. We were able to gather a great deal of geo-tagged tweets from

Twitter in the geographic range of Japan. We computed the aforementioned crowd behavioral features based on three parameters using metadata of geo-tagged tweets and extracted urban characteristics by examining their common changing patterns. Lastly, we successfully confirmed that our experimentation to utilize the social network based crowd behavior as an estimator for urban characterization was achieved by investigating categories of major facilities in each clustered region.

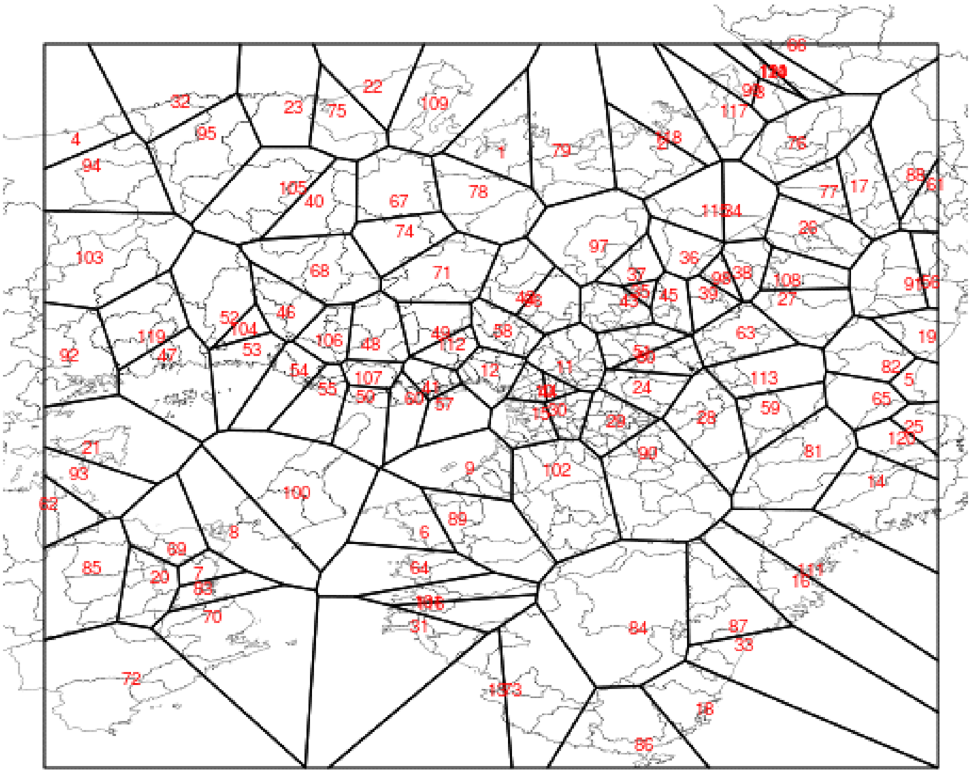
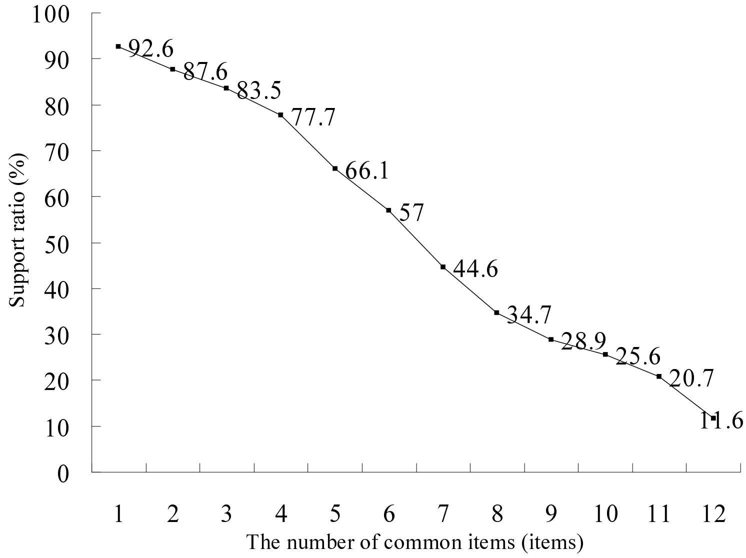


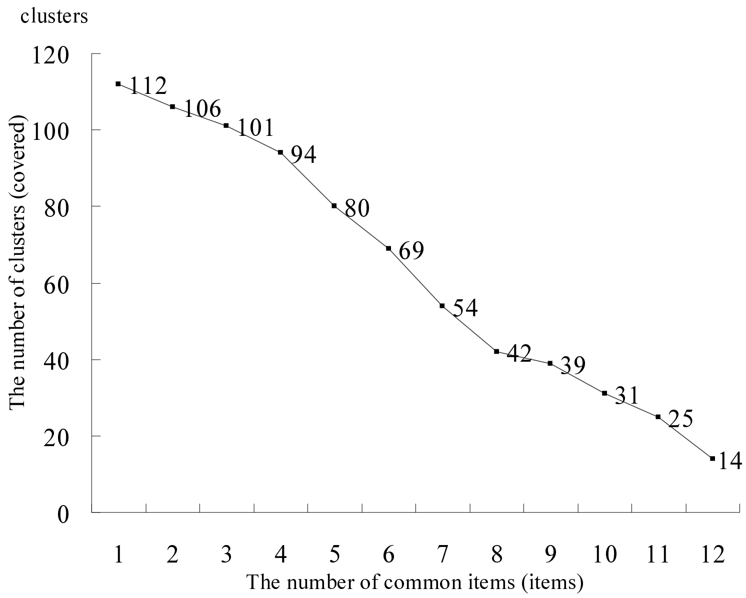
Figure 27: Urban Areas Partitioned by Socio-geographic Boundaries

5.4.1 Experimental Setting

First, we collected geo-tagged tweets from Twitter for one month between Jun. 5th, 2010 and Jul. 5th, 2010 around part of Japan with the latitude range [33.384555:35.839419] and the longitude range [134.126551:136.58890] using our geographical microblog monitoring system as shown in Figure 23 (a). As a result, we could acquire 1,891,186 geo-tagged tweets from 39,898 distinct users as shown in Figure 23 (b). Next, we constructed socio-geographic boundaries considering the density of crowd behavior and partitioned the target space into 121 urban clusters as shown in Figure 27.



(a) Relation between support ratio and the number of common items



(b) Relation between the number of clusters and the one of common items

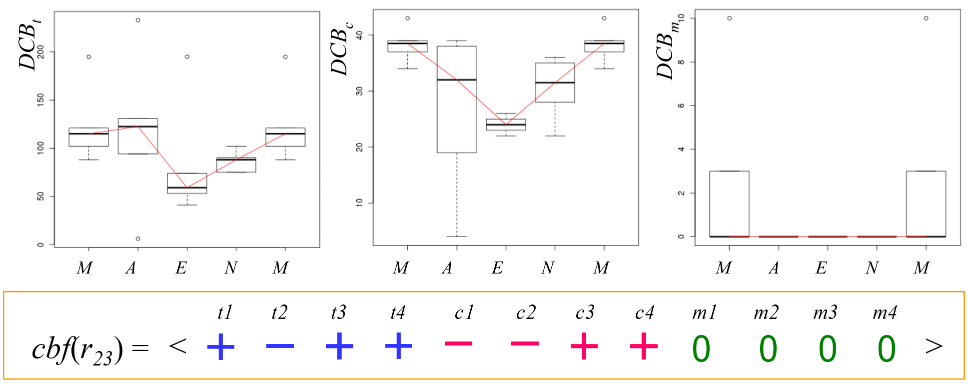
Figure 28: Effects of Common Item Control on Support and Covered Clusters

5.4.2 Exploring Significant Crowd Behavioral Patterns

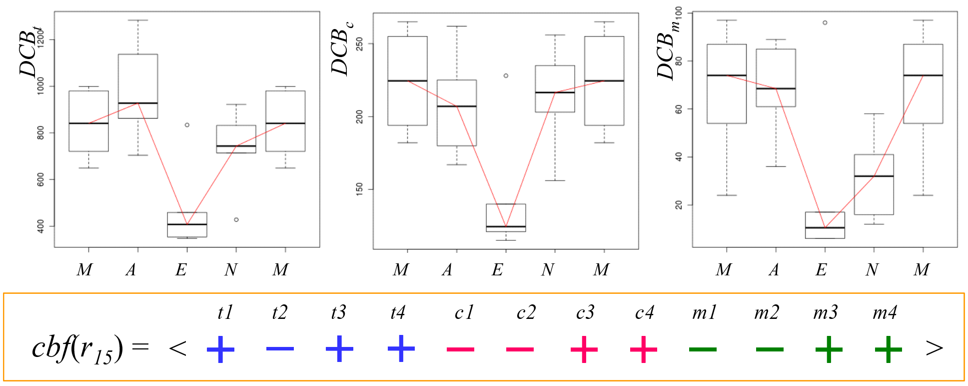
Next, in each cluster, we computed crowd behavioral feature based on three parameters, for every 6-h time slot during the training period. Then, we extracted significant crowd behavioral patterns by analyzing the crowd behavior features for all clusters. In this experiment, we applied a frequent itemset mining algorithm [88] which can statistically mine common features to full size: 12 items. Specifically, we tried to extract significant features by examining partial patterns from the 12-dimensional crowd behavioral features. By applying the pattern extraction method, we were able to extract characteristic and frequent crowd behavioral patterns. Here, we can consider two important parameters for the pattern extraction. First, each pattern can have a support value which says how many clusters actually support the correspondent patterns. By specifying the support value, we can control the resulting number of patterns. Furthermore, we can consider the size of pattern, that is, how many number of each cluster should meet the patterns. For instance, for a pattern $< t1+$, $t2*$, $t3+$, $t4-$, $c1*$, $c2+$, $c3*$, $c4-$, $m10$, $m2+$, $m3-$, $m4- >$, three parts of '$*$' symbol can include any cases of '+,' '-' or '0,' we also specified the preferable size of patterns from 12 down to only 1. Consequently, the effect of the support value and the common item size was examined as illustrated in Figure 28 (a). Another important aspect is that depending on the size of common items, the number of clusters covered by the extracted patterns is decided as drawn in Figure 28 (b). In other words, when we set the size of pattern shorter, many clusters can be included in the final results. Finally, with a setting with the preferable size of frequent itemset (=12) and minimum support (=3.0%), we could obtain 4 significant crowd behavioral patterns as depicted in Table 7. Among these patterns, we examined 2 interesting patterns of $pattern1$ and $pattern3$ as shown in Figure 29 (a) and (b), respectively. The reason why we select $pattern1$ is that the occurrence ratio of the pattern is highest and covers about 14 urban clusters. Then, the reason of selection of $pattern3$ is that it includes increase and decrease of the number of movings.

Table 7: Crowd Behavioral Patterns Extracted based on the Frequent Itemset Mining Algorithm (12 items)

pattern	DCB_t				DCB_c				DCB_m				occurrence ratio (%)
	$t1$	$t2$	$t3$	$t4$	$c1$	$c2$	$c3$	$c4$	$m1$	$m2$	$m3$	$m4$	
$pattern1$	+	−	+	+	−	−	+	+	0	0	0	0	11.6
$pattern2$	−	−	+	+	−	−	+	+	0	0	0	0	7.4
$pattern3$	+	−	+	+	−	−	+	+	−	−	+	+	5.0
$pattern4$	+	−	+	+	+	−	+	+	0	0	0	0	3.3



(a) A cluster of *pattern 1*



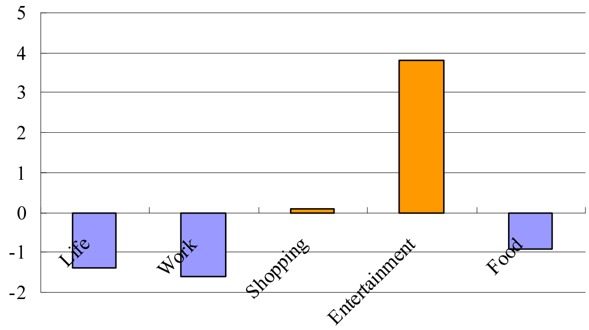
(b) A cluster of *pattern 3*

Figure 29: Examples of Crowd Behavioral Patterns

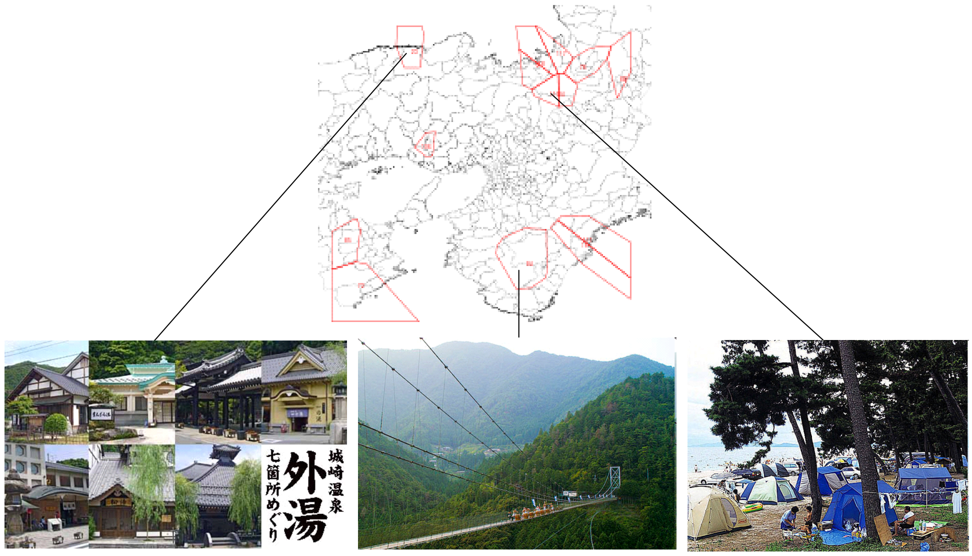
5.4.3 Reasoning Urban Characteristics

In order to reason the extracted patterns and characterize urban clusters of the same patterns respectively, we investigated what kinds of local facilities can be found in the clusters respectively. In order to obtain a baseline dataset, we prepared a database of local facilities by referring to geographic facilities information of the local search provided by Yahoo! Japan (Yahoo! Loco) [89] which answers surrounding facilities' information with precise location information in Japan. Most facilities have a hierarchy of three levels, respectively, for instance, 'Food' (1st), 'Chinese' (2nd), 'Beijing Food' (3rd) and can be eventually aggregated into four genres at the most upper level; 'Food,' 'Shopping,' 'Entertainment,' and 'Life.' However, there were some facilities that are not allocated yet to any category above such as factory and distribution center. We found that these facilities are mostly relevant to industrials, thus we manually added a new genre 'Work' and sub-categories by referring to Japan Standard Industry Classification [90].

Based on this local facilities database, we looked into the significant facilities of each region in terms of five genres; 'Life,' 'Work,' 'Shopping,' 'Entertainment,' and 'Food' as shown in Figures 30 (a) and 31 (a). Here, the Y-axis in the graphs of Figures 30 (a) and 31 (a) represents a relative significance compared to the average over all clusters. In other words, if the value of each genre is under zero, it represents that the genre is weaker than the average. For example, the clusters grouped by *pattern*1 are commonly characterized by the genre feature as shown in Figure 30 (a). That is, the type of regions relatively have many entertainment facilities such as hot springs, camping parks, sightseeing spots, etc. In fact, we can find entertainment facilities in urban clusters of *pattern*1 as shown in Figure 30 (b). On the other hand, the clusters grouped by *pattern*3 are characterized as relatively many life and work facilities such as residences, educational facilities, and office buildings as illustrated in Figures 31 (a). We can actually confirm residential towns, universities, and high-rise buildings for business in urban clusters of *pattern*3 as shown in Figure 31 (b). In fact, each cluster includes a variety of facilities eventually showing mixed crowd behavior. In other words, it is hard to say that a town only has a specific characteristic such as only shopping town. In our work, instead, we were able to find out the distinctive and significant configurations of local facilities in the regions clustered by crowd behavioral patterns to understand the characteristics of urban regions.



(a) Facility genre-based reasoning of crowd behavioral *pattern 1*



Hot springs
Kinosaki onsen, Hyogo

Sightseeing spots
Totsugawa, Nara

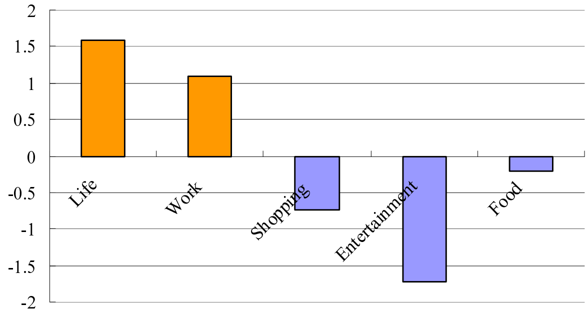
Camping facilities
Biwa-lake, Shiga

(b) 14 clusters of crowd behavioral *pattern 1*

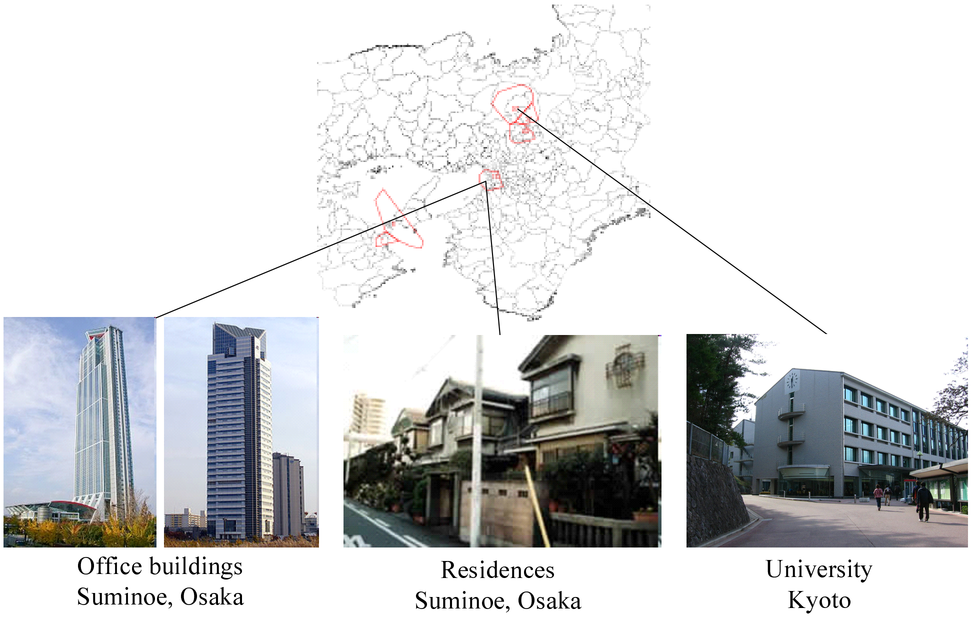
Figure 30: Reasoning Characteristic of Urban Clusters of $pattern1$

5.5 Summary

In this paper, we proposed a crowd-based urban characterization method based on crowd-sourced data obtained from social networking sites. In our proposed method, we collected geo-tagged tweets from Twitter as a source of monitoring crowd behavior in the real world and modeled crowd behavior in terms of primitive parameters extracted from the tweets. Then, we extracted crowd behavioral features for urban areas and investigated urban characteristics by exploiting common behavioral patterns. Furthermore, on the basis of our proposed method, we conducted the experiment using 1,891,186 geo-tagged tweets



(a) Facility genre-based reasoning of crowd behavioral *pattern 3*



(b) 6 clusters of crowd behavioral *pattern 3*

Figure 31: Reasoning Characteristic of Urban Clusters of *pattern*3

from 39,898 users between Jun. 5, 2010 and Jul. 5, 2010 in Kinki area and clustered urban areas based on 4 crowd behavioral patterns which are significantly extracted. Finally, in order to reason each grouped urban areas, we investigated genre information of local facilities in the urban areas.

Chapter 6 Crowd-sourced Cartography: Measuring Socio-cognitive Distance for Urban Areas based on Crowd's Movement

6.1 Introduction

Urban space is a complicated mixture, which includes a variety of elements; physical objects such as local facilities and landmarks, natural phenomena like climates and disasters, and social activities such as cultural or political events, etc. In such a complex space, we are always required to conduct various location-based decision makings from looking for a restaurant for daily lunch to exploring a new dwelling. In such situations, final decisions would be often elicited depending on individual experiences in an urban area or limited knowledge about the area; if a person frequently visits a city, s/he intuitively thinks that the city is more familiar than other cities where s/he has been less. Therefore, based on personal experiences to a sophisticated urban structure, we become to have a bound image of the urban space and eventually make an unsatisfactory choice. Here, we regard such individually distorted image of urban areas as a cognitively recognized urban space. For instance, let's assume a situation where a person is looking for a place to live with his family. He would like to find an ideal place which can meet various demands; not only accessibility to his workplace in terms of public transportation, but also convenience for shopping, safety, cleanliness, educational environment of the place, etc. For this, he may first consult a general reference map for making a short list of candidate places to live by considering the accessibility based on transportation convenience. General reference maps like Google Maps [91] would be the first step to look up general features of urban space; cities, roads, railways, local facilities, etc. Especially, a travel time map or commute map can show the adjacent neighborhood of the workplace in terms of not only geographical proximity but also the accessibility based on transportation convenience in the urban space. However, such a map just shows a commutable area which includes lots of candidate places to move. Therefore, in order to find out complex and dynamic local characteristics, general reference maps are not enough to provide appropriate answers for much sophisticated questions such as “Which place can give better educational environment for his children?” Hence, in order to examine this kind of local characteristics or knowledge, we further need to search for the Web, variously local statistics information by public administration, word-of-mouth from acquaintances, etc. However, it is not easy to acquire the local characteristics without huge costs and efforts as illustrated in Figure 32 (a).

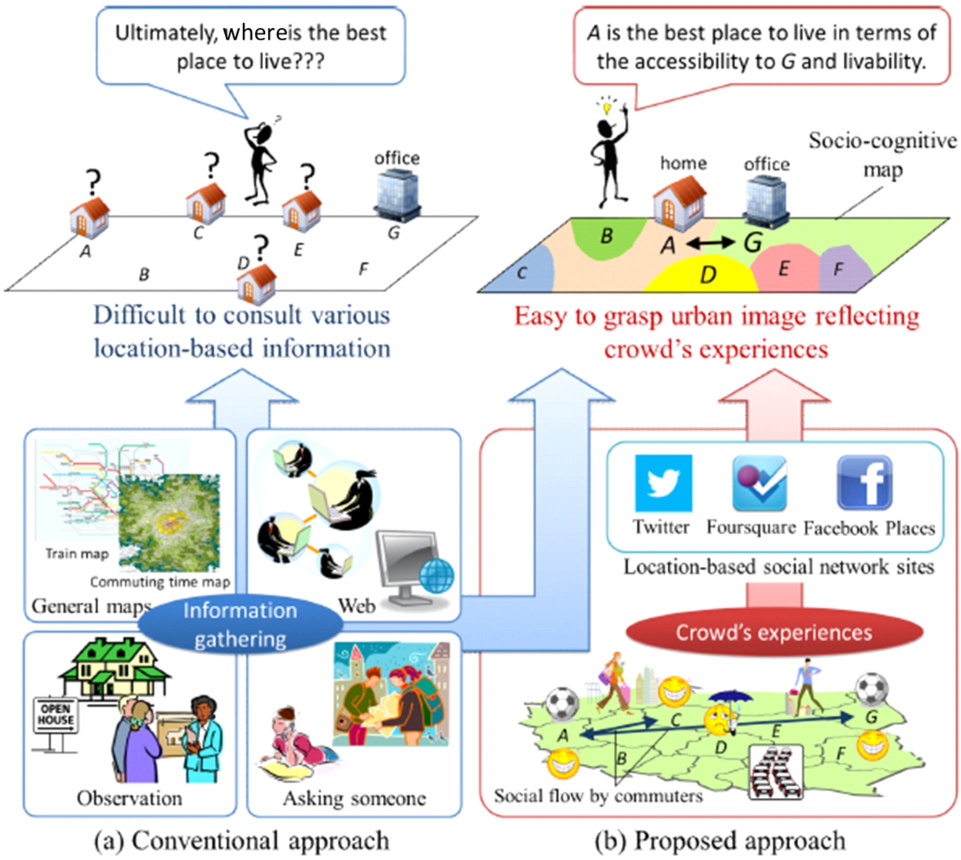


Figure 32: Motivation for Socio-cognitive Map Generation

In particular, this paper will focus on the novel crowd-sourced lifelogs which are represented by Twitter; we will explore new values of the massive urban people’s experiences as a source to explore various urban dynamics and characteristics relevant to crowd’s lifestyles utilizing the recent location-based social network sites such as Twitter [3], Foursquare [5], and Facebook Places [92]. Furthermore, with the extracted crowd's lifestyles in an urban area, we aim to generate an integrated map which can represent various local characteristics in an urban space as shown in Figure 32 (b). Particularly, we focus on the accessibility by measuring socio-cognitive proximity based on crowd's movements in an urban space.

Furthermore, we present a method to generate a socio-cognitive map as a kind of thematic map based on Twitter-based crowd's movements. Obviously, cartography based on crowd-sourced lifelogs would be an interesting and important challenge to provide much useful local information. Generally, in a specialized map called cartogram [93], mapping variables such as travel time, population or incomes are substituted for land area or distance; the geometry or space of the map would be significantly distorted to represent the information of this alternate variable. For the crowd-sourced cartography, we first look for significant places as social urban clusters in an urban space based on the density of crowd through location-based social networks. Next, we measure influential strength of each social urban cluster for an area variance. Then, we compute socio-cognitive distance between the clusters based on the crowd's movements for a distance variance. Finally, we generate a socio-cognitive map by projecting urban clusters on two-dimensional space by means of MDS (Multi-Dimensional Scaling) as well as by emphasizing urban clusters based on their influential strengths with a Weighted Voronoi Diagram. The contributions of this work are summarized as follows.

- We generated a socio-cognitive map by exploiting crowd-sourced lifelogs on location-based social networks.
- We defined crowd-sourced cognitive distance by expanding the concept of cognitive distance.
- We developed a technical method to generate socio-cognitive cartogram.

6.2 Computing Social Urban Structure through Location-based Social Network

In this work, in order to represent complex and dynamic urban space, we attempt to generate a socio-cognitive map of urban space by exploiting local crowd's experiences. For this, we utilize crowd's lifelogs publicly shared on recent location-based social network sites. In order to monitor crowd's experiences using social network sites, we modeled crowd's experiential features. In general, lifelogs on most location-based social network sites are consisted of a set of metadata such as user ID, timestamp, location information and textual message. In case of Twitter, we can extract further useful metadata such as reply words in a textual message, hashtags, followers or following relationships, retweets, links to external media, etc. On the basis of such metadata, we can first define personal experiential features consisting of five indicators relevant to individual experience; 1) user's existence in an urban cluster which is represented by user ID and location information, 2) user's activity in terms of publishing tweets and moving in an urban cluster which is computed by using user ID, timestamp and location information, 3) user's sentiment [51] which can be computed by determining sentimental words or the ratio of positive or negative words in a textual message, 4) user's interest which is represented based on textual hints such as topic keywords and hashtags as well as links to external media like Web pages, photos, video clips, etc., and 5) user's relationships and interactions with other users which are computed based on followers or following relationships, replies and retweets.

Next, on the basis of the five types of indicators above constituting personal experiential features, we can define crowd experiential features as shown in Figure 33. In sum, by aggregating individual existences in an urban cluster, we can easily grasp demographics such as crowd's population and density populated points in the urban cluster. Individual activities can represent crowd's activity and movements in terms of congestion and activation. Users' sentiments show a mood in an urban cluster. For instance, when lots of people in an urban cluster relatively feel happier than other clusters, we can expect that the mood of the urban cluster reflected on the location-based social networks would be also positive. In addition, the aggregated personal interests could be regarded as crowd's topics and social trends. Crowd's social networks are clearly connected with social communications and we can grasp socio-geographical relationships among urban clusters through human relationships. In this paper, we will measure a socio-cognitive distance based on accessibility between urban clusters by especially focusing on crowd's movements as one of the indicators.

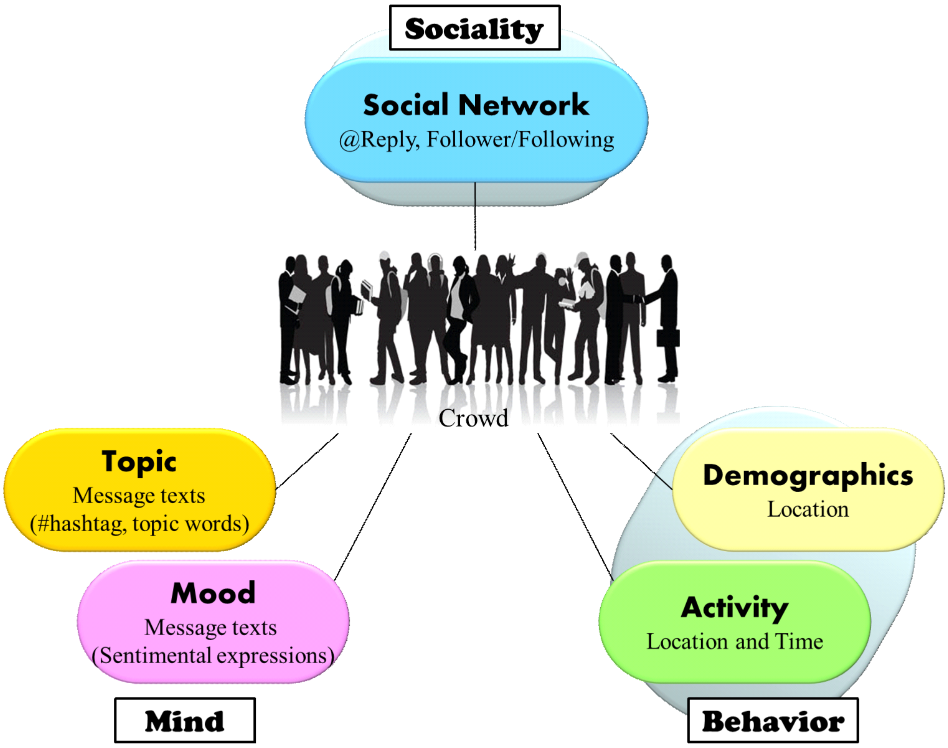


Figure 33: Crowd Experiential Features Extractable from Location-based Social Networks

6.3 Generating a Socio-cognitive Urban Map

In this subsection, we describe our map generation method in an order as shown in Figure 35, which begins with collecting crowd's lifelogs from location-based social network. Finally, we generate a map of our interest which aims to deliver an intuitive urban structure focusing on practical proximity between urban clusters; that is, accessibility, by investigating crowd's movements.

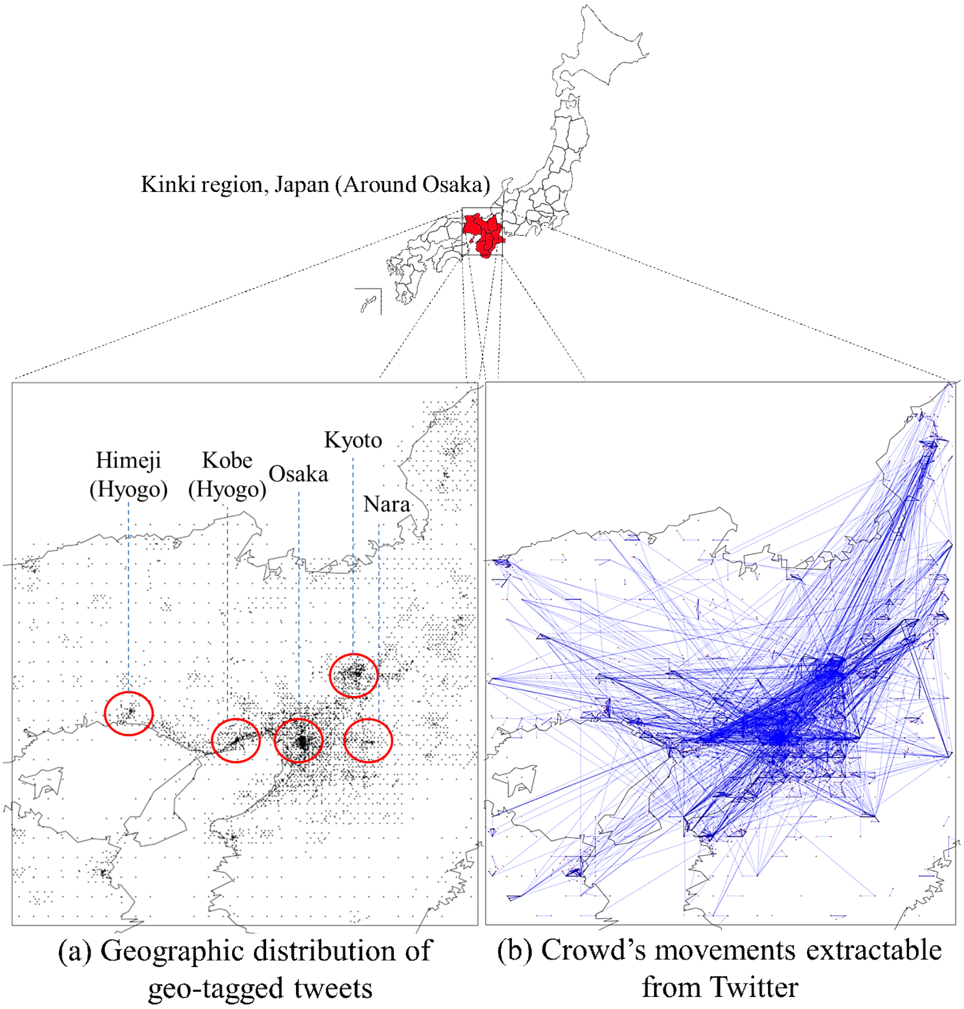


Figure 34: Geographic Distribution of Geo-tagged Tweets and Crowd's Movements Monitored through Twitter

6.3.1 Collecting Crowd's Movements from Twitter

We first gather geo-tagged tweets from Twitter to monitor crowd's movements in an urban space as shown in Figure 35 (1). However, it takes a considerable amount of efforts to acquire a significant number of geo-tagged tweets due to practical limitation of an open API provided by Twitter which only supports the simplest near-by search based on a specified center location and a radius and obtains a limited number of tweets. In order to overcome this problem, we developed a geographic tweets gathering system, in our previous work [42], which can monitor crowd behavior for a specific region of any size depending on

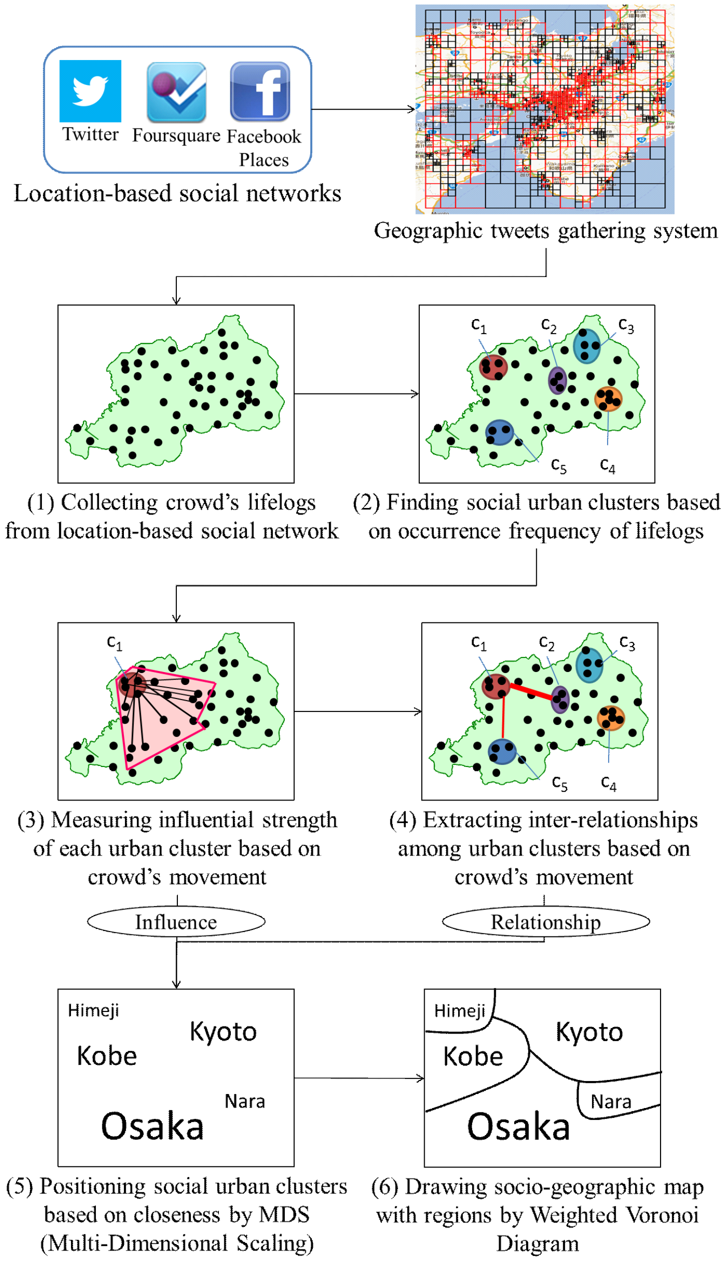


Figure 35: Procedure for Generating Cognitive Map based on Crowd's Movements

the density of massive geographic microblogs for overcoming these limitations and carry on monitoring of any size of user-specified regions.

Based on the geo-tagged tweets, we extract crowd's moving segments by exploiting primary metadata of geo-tagged tweet; user ID, timestamp, and location information. After we determine when and where each tweet was written, we can plot their location points on the map as shown in Figure 34 (a). Furthermore, in order to effectively search moving segments with the unified dataset, we assemble the tweets based on user ID and sort each user's tweets in the order of the timestamp. Consequently, Figure 34 (b) shows moving segments of crowds observed through Twitter.

6.3.2 Locating Urban Clusters

Most thematic maps are generated by emphasizing some landmarks or characteristic areas according to various purposes of cartographers respectively. We will also generate such kind of cognitive map, where some geographic features appeared on the maps are appropriately selected. For this, we locate social urban clusters by utilizing the density of crowd as shown in Figure—35 (2).

However, in this paper, we are based on massive number of crowd's lifelogs found on Twitter, hence, it would require unbearable computational efforts to find out social urban clusters. Therefore, we need to reduce the data size in much smaller and compact size without loss of essential quality of the original data. For this, we adopted the NNClean algorithm [83] to split the data into two classes of high-frequency and low-frequency parts.

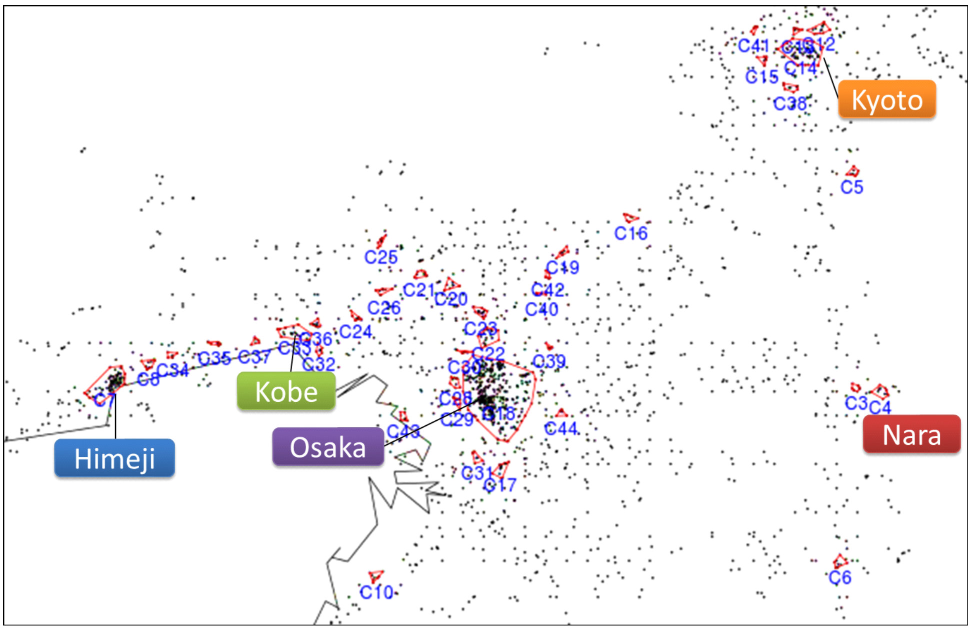


Figure 36: Urban Clusters Generated based on Crowd's Lifestyles

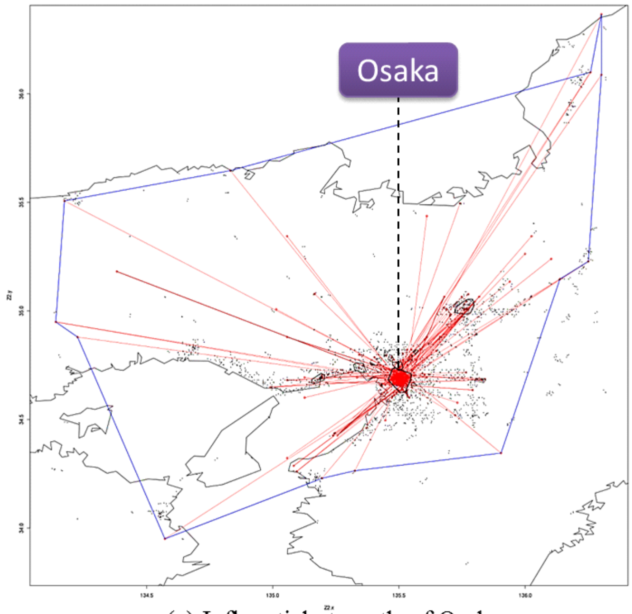
Then, we look for social urban clusters with the ideally reduced dataset. In order to find high-density areas, we apply one of conventional clustering methods; DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [94]. The benefit of this algorithm is the ability to deal with non-Gaussian distributed data because it features a cluster model called density-reachability. This algorithm connects points that satisfy a density criterion, in the original variant defined as a minimum number of points MinPts within a specified radius. A cluster consists of all density-connected points which can form a cluster of an arbitrary shape. In the experiment, we generated a socio-cognitive map with 44 urban clusters located by empirically setting $MinPts$ and $radius$ to 6 and 0.0065, respectively. Then, we eventually represented the clusters by convex-hull based boundary polygons [95] which center points in a bounding convex as shown in Figure 36.

6.3.3 Measuring Influential Strength of an Urban Cluster

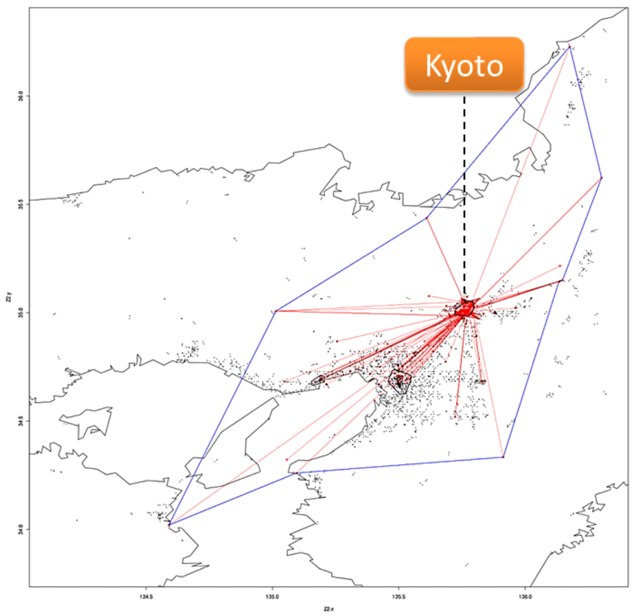
Next, by examining where people come and go out to an urban cluster, we can know its influential strength in an urban space as shown in Figure 35 (3). As shown in Figure 37, it would be possible to assume that the influential strength is the overall area including all of connected places with the urban cluster. The measured influential strength of each social urban cluster is utilized when visualizing a socio-cognitive map. Here, we illustrate Figure 37 (a) and (b) which show urban clusters where influences are the maximum in our finding. Its effective ranges cover the most major landmark areas in our experimental area of Kinki region in Japan (including Osaka, Kyoto and Kobe). From this result, we confirmed that the influential strength of an urban cluster can be simply computed by the total number of moving segments.

6.3.4 Calculating Cognitive Distance between Urban Clusters

Then, we measure cognitive distances between urban clusters respectively as shown in Figure 35 (4). We assume that if the shorter the physical distance between urban clusters is and the more frequently crowds move the clusters, then the closer the clusters would be regarded by people. Therefore, we consider that urban clusters which are closely associated with each other should be projected closely on a cognitive map. On the basis of the hypothesis, we made



(a) Influential strength of Osaka



(b) Influential strength of Kyoto

Figure 37: Influential Strengths of Urban Clusters

a formula to calculate a cognitive distance between urban clusters as follows:

$$CogDist(c_i, c_j) = w_1 \cdot EucDist(c_i, c_j) + w_2 \cdot ExpDist(c_i, c_j) \tag{13}$$
$$(w_1 + w_2 = 1.0, w_1, w_2 \geq 0)$$
$$ExpDist(c_i, c_j) = \frac{1}{\#MovSeg(c_i, c_j) + 1} \tag{14}$$

where three functions, *CogDist*, *EucDist* and *ExpDist*, calculate distances between urban clusters (c_i, c_j) in terms of cognitive, physical, and experiential, respectively. Specifically, the function *CogDist* is calculated by *EucDist* and *ExpDist*,. The function *EucDist* calculates normalized Euclid distance between urban clusters, and the function *ExpDist* calculates normalized experiential distance between them based on the quantity of crowd's movements given by a function $\#MovSeg$ which counts the number of moving segments between the clusters. The values computed by *EucDist* and *ExpDist* are weighted based on given values, w_1 and w_2, respectively. The weight values can be freely set on a user's purpose for generating a cognitive map. For example, if a user wants to generate a cognitive map by emphasizing on the crowd's movements, s/he can set a high weight to w_2. In the experiment, we show cognitive maps generated with different pairs of the weight values.

6.3.5 Projecting Closeness between Urban Clusters

Next, we plot the computed closeness among urban clusters based on crowd's movements as shown in Figure 35 (5). In order to intuitively represent a socio-cognitive distance of urban clusters by measuring closeness cognitively recognized among them, we need to appropriately allocate the clusters on a socio-cognitive map. In this paper, as shown in Figure 38, we decided to apply Multi-Dimensional Scaling (MDS) [96], which allocates given dataset of multi-dimensional space into a low-dimensional space by considering similarities or dissimilarities in the dataset. In sum, it can allocate two urban clusters in the neighborhood if the similarity between them is high. In contrast, if the similarity between them is low, it allocates them far away. Specifically, the MDS algorithm starts with a matrix of item-item similarities, then assigns a location to each item in N-dimensional space, where N is specified a priori. In the experiment, we mapped social urban clusters in a two-dimensional space. We also

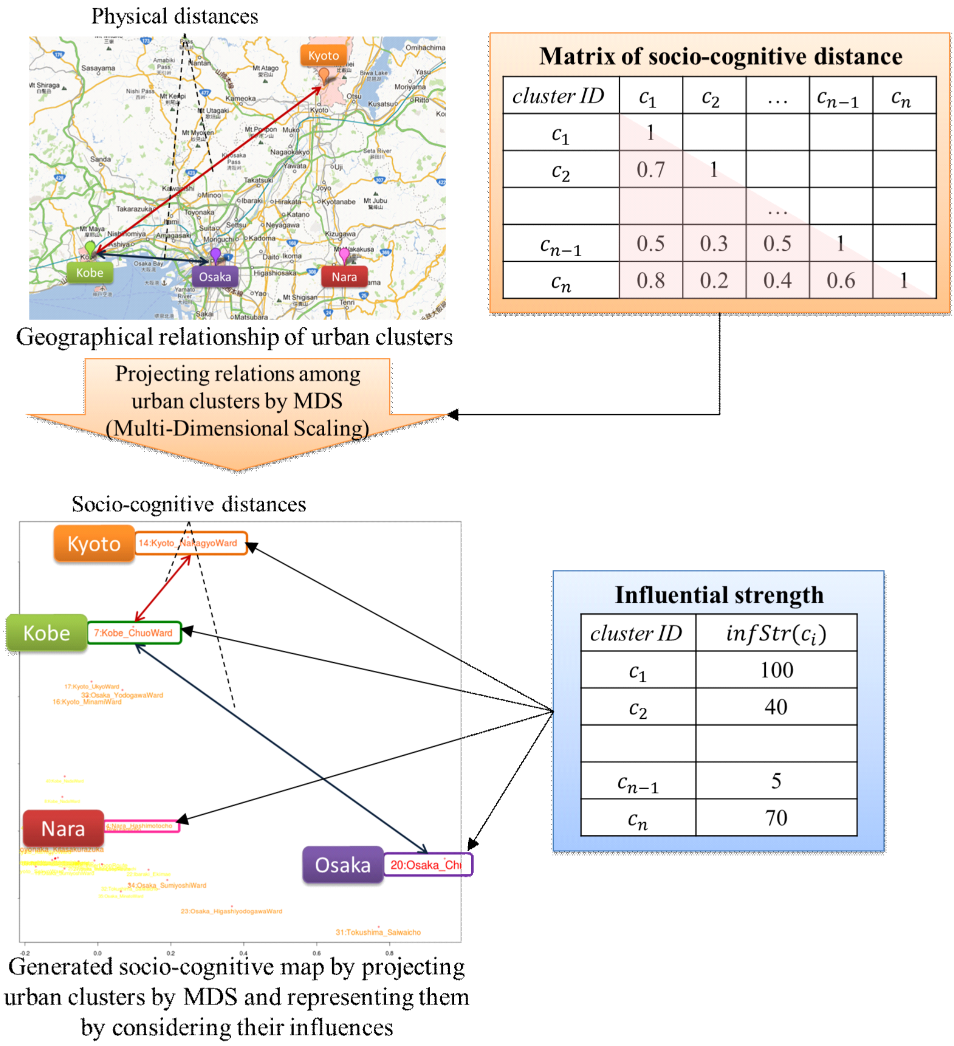


Figure 38: Projection of Urban Clusters in terms of Area and Distance based on MDS

labeled names of generated clusters by a Reverse Geocoding service [82] which can translate a given location coordinate into a textual place name. For each cluster, we obtain its representative place name by using this service.

6.3.6 Drawing Socio-cognitive Regions

In previous subsection, we plotted social urban clusters which are labeled by approximate geographic names. However, it is hard to regard the result as a cognitive map because the representation can just show socio-cognitive closeness of urban clusters. Urban clusters on a socio-cognitive map would be better allocated with region-based space partitioning as shown in Figure 35 (6). For

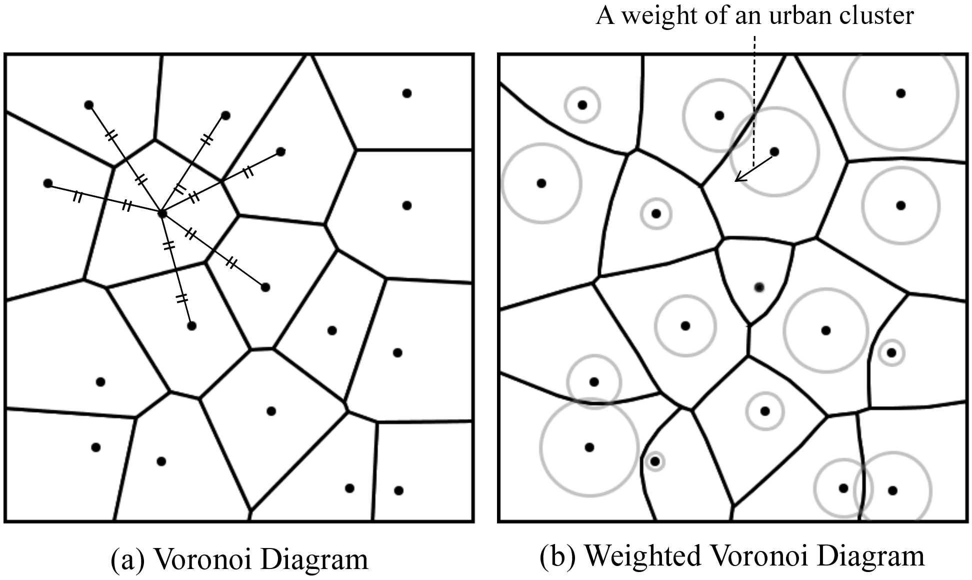


Figure 39: Examples of Voronoi Diagrams

the purpose, we applied a Weighted Voronoi Diagram [97] to the result of MDS. Generally, a Voronoi Diagram depicted in Figure 39 (a) is known as an algorithm for partitioning a space by drawing a line between two points keeping with the same distance. However, urban clusters do not have the same influential strength; we can solve this problem by means of a Weighted Voronoi Diagram, where each cell can have a weight extending its size compared to normal Voronoi Diagram as shown in Figure 39 (b).

6.4 Experiment

In this section, we describe our experiment to generate a socio-cognitive map. For this, we collect massive geo-tagged tweets for a day from Twitter in an area of Japan. Then, we located social urban clusters based on the geographic distribution of crowd's lifelogs and measured an influential strength of each urban cluster based on crowd's moving segments. Next, we computed cognitive relationships among the clusters in terms of physical and social experiential distances. Finally, we generated a socio-cognitive map based on the extracted crowd's moving patterns.

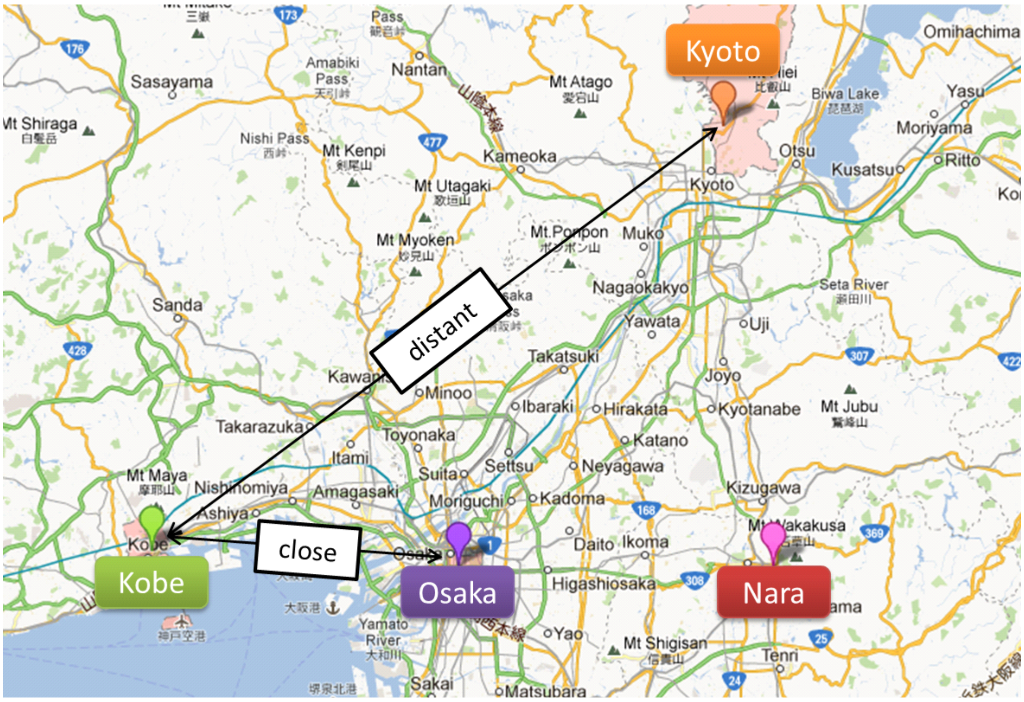


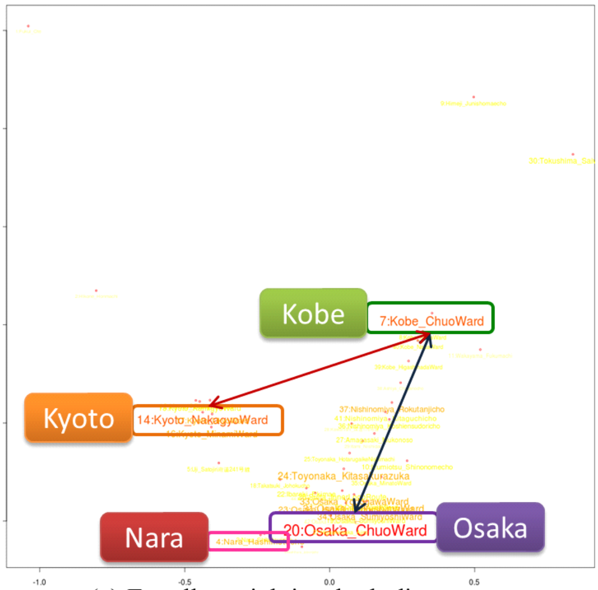
Figure 40: Positional Relationship of Urban Clusters

6.4.1 Dataset

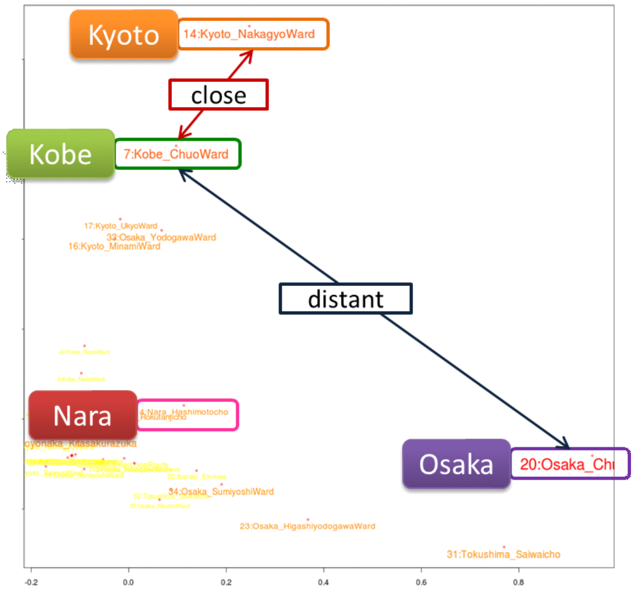
We collected 157,097 geo-tagged tweets from 25,674 distinct users in a day, April 23rd, 2012 in a surrounding area including Kobe, Osaka, Kyoto and Nara in Japan (longitude range = [134.122433, 136.337186], latitude range = [33.810804, 36.785050]) from Twitter using the geographical tweet gathering system [42]. From each geo-tagged tweet, we utilized spatio-temporal clues; user ID, timestamp, and location coordinate for monitoring crowd's movements in an urban space.

6.4.2 Generated Socio-cognitive Map

We first located convex-hull based urban clusters by applying DBSCAN algorithm with location points of the dataset reduced by NNClean method as shown in Figure 36. Next, we determined the influential strengths of the urban clusters for showing their relations by computing crowd's moving segments. Figure 37 shows examples of urban clusters measured strong influence on wide surrounding areas. From this result, we confirmed that we could capture the influential strength of each urban cluster in a simple way through the quantity of crowd's moving segments relevant to each urban cluster.



(a) Equally weighting both distances
(w_1= 0.5, w_2= 0.5)

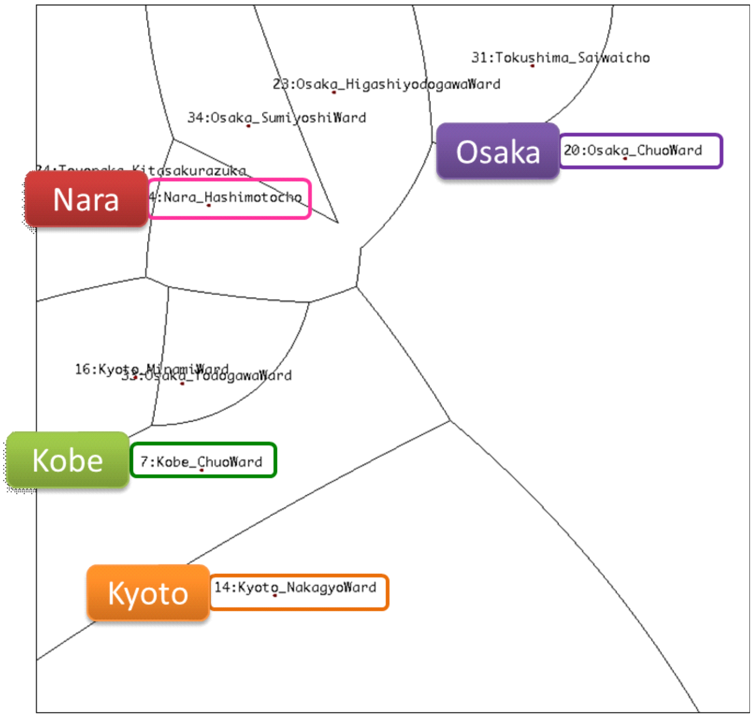


(b) Weighting the experiential distance
(w_1= 0, w_2= 1.0)

Figure 41: Social Urban Clusters Projected by MDS

Then, we calculated socio-cognitive distances between urban clusters for counting crowd's moving segments between the clusters and projected the clusters with their relationships to socio-cognitive maps by applying the MDS algorithm. Here, we can freely set weight values for physical distance and experiential distance which bringing the total to 1.0. As aforementioned, a user can easily adjust the degree of his/her requirements. We show the plotted urban clusters which set different weight values in the formula (14) by MDS algorithm as illustrated in Figure 41. Texts appeared on a two-dimensional space are local address at the center of each urban cluster, and their sizes and colors mean their influential strengths; the bigger the size of a cluster's label is and the darker its color is, the more influential the cluster is regarded. In detail, Figure 41 (a) shows a result based on socio-cognitive distance measured with the same weight (= 0.5) to both physical and experiential distances among urban clusters. In this case, geographically close urban clusters were relatively aggregated closely after the MDS computation. In contrast, Figure 41 (b) is the result by attaching a high weight (= 1.0) to experiential distance based on crowd's moving segments between urban clusters. Thus, we can grasp localized social relationships reflecting crowd's movements. Furthermore, in both figures, we interestingly found that socio-cognitive distances relevant to Kobe are different from the physical distances in the real world as shown in Figure 9; the physical distance between Kobe and Osaka is closer than the one between Kobe and Kyoto as shown in Figure 40, but the socio-cognitive distance between Kobe and Osaka is measured more distant than the one between Kobe and Kyoto as shown in Figure 41 (b).

Next, we allocated regions for each urban cluster just plotted on the two-dimensional space. As shown in Figure 42, we can successfully generate a socio-cognitive map consisting of the top 10 urban clusters ranked by their influential strengths by means of a Weighted Voronoi Diagram. As shown in Figure 42, we can obtain a socio-cognitive map having regions of urban clusters in terms of crowd movements. Here, we can find two influential urban clusters; ChuoWard, Osaka (cluster ID is 20) and NakagyoWard, Kyoto (cluster ID is 14). Subsequently, we explain an interesting result, which is different from the real space. For example, regions of clusters in Osaka mostly covered a large part on this map. As described above, ChuoWard, Osaka (cluster ID is 20) and NakagyoWard, Kyoto (cluster ID is 14) are close to each other. In addi-



Socio-cognitive map generated by weighting the experiential distance (w_1= 0, w_2= 1.0)

Figure 42: Generated a Socio-cognitive Map

tion, there are other clusters in Osaka such as SumiyoshiWard (cluster ID is 34), YodogawaWard (cluster ID is 33), etc. are close to the cluster in Nara, Hashimotocho (cluster ID is 4). Finally, we were able to take advantages of the representation based on the Weighted Voronoi Diagram which can help us understand more intuitive and simplified understanding among urban clusters. As for measurement of socio-cognitive proximity, it depends on characteristics of crowd movements extracted from microblogs, that is, although the proximity between Kobe and Kyoto was measured as closer than the proximity between Kobe and Osaka, people who go to Kyoto is generally go through Osaka. Like this, the observed crowd movements would be different from the real crowd movements. However, by clarifying and considering characteristics of the crowd movements from microblogs, we will address the difference and develop various applications based on our proposed proximity. Furthermore, we will conduct on meaningful socio-cognitive map generation by further observing crowd's experiences extractable from social networks.

6.5 Summary

In this section, we proposed a method to generate socio-cognitive map by measuring influential strength of each urban area and examining social relationships among social urban areas based on crowd's movements monitored by exploiting geo-tagged tweets over Twitter. Specifically, we collected $157,097$ geo-tagged tweets published from 25,674 users on April 23, 2012 (a weekday) in Kinki area and measured socio-cognitive distance among urban clusters based on crowd movements. Finally, we could show the possibility to represent socio-cognitive distances between urban clusters and influential strengths of urban clusters on a map.

Chapter 7 Discussion

In this doctoral dissertation, we introduced four approaches for exploring and utilizing crowd-sourced contents collected from social media in terms of collaborative and spatio-temporal aspects.

Cooperative Analysis As one of noteworthy characteristics of social media, we focused on cooperative aspect and made two approaches; (a) Crowd-powered Scene Extraction for Shared Video Clips and (b) Crowd-powered Video Rating. In the case (a) described in Section 3, we analyzed relations of pointing regions and temporal durations of user comments which are attached to shared video clips. Then, we found and utilized the relations for extracting relevant scenes in terms of objects and events. In the case (b) described in Section 4, we analyzed microblogs written by massive users to extract crowd's media consumptions to TV programs or on-line videos and attempted to utilize the found microblogs for conducting a comprehensive video media rating including TV programs and on-line videos. In the experiment, we measured relevance between $119,575$ geo-tagged tweets collected from Twitter and $838,636$ EPG items ($24,841$ distinct TV programs) from an EPG site or several keywords which are able to identify videos. The two approaches we made for studying the advanced video media can have further mutually beneficial relationship as shown in Figure 43. In order to put our research work to practical use, we consider future work particularly based on the emerging synergetic relationships as follows.

- By extracting microblogs relevant to on-line video clips shared over social media services, we will be able to obtain much more hints such keywords, significance of attention by crowds, eventually useful for extracting relevant scenes on video clips as depicted in Figure 43 (a). Consequently, it is expected to improve the accuracy of object and event determination methods, respectively, which we approached in the case (a).
- Based on relations of user comments for video clips measured in the case a), we will be able to compute local popularity of each video clip

as illustrated in Figure 43 (b). By utilizing this popularity for the case (b), we can extend rating of videos in much finer granularity such as object, event, scenes, and videos.

Spatio-temporal Analysis Focusing on the real-life uses of social media, we introduced two approaches on spatio-temporal analysis; (c) Crowd-sourced Urban Area Characterization and (d) Crowd-sourced Cartography. In the case (c) explained in Section 5, we established a method for characterizing urban clusters grouped by extracting significant crowd behavioral patterns. In the experiment, we collected $1,891,186$ geo-tagged tweets from Twitter which were written by $39,898$ distinct users between Jun. 5, 2010 and Jul. 5, 2010 in Kinki area and extracted 4 significant crowd behavioral patterns by analyzing crowd behavior features monitored within each urban cluster. Then, in order to interpret characteristics of urban clusters grouped based on the patterns, we observed the categorical distribution of the grouped urban clusters, respectively. In the case (d) described in Section 6, we proposed a method to generate a map by measuring socio-cognitive distance between urban clusters and influential strengths of urban clusters based on crowd movements. We collected $157,097$ geo-tagged tweets published from $25,674$ distinct users on April 23, 2012 (a weekday) in Kinki area and measured socio-geographic relations between urban clusters based on crowd movements. Then, we projected urban clusters based on the relations on a map.

Considering that spatio-temporal analyses in these two cases can be integrated as illustrated in Figure 43, we will show future direction based on our study experiences as follows.

- Urban characteristics would be examined by crowd behavior features such as connections between remote urban clusters by massive commuters. For instance, the relation between urban clusters based on the quantity of crowd movements analyzed in the case (d) will be utilized as a crucial and useful parameter of crowd behavioral features as shown in Figure 43 (c).
- In order to measure a sense of proximity between the urban clusters, we

can also approach it with characteristics of the urban clusters, which can be extracted in (c) as shown in Figure 43 (d).

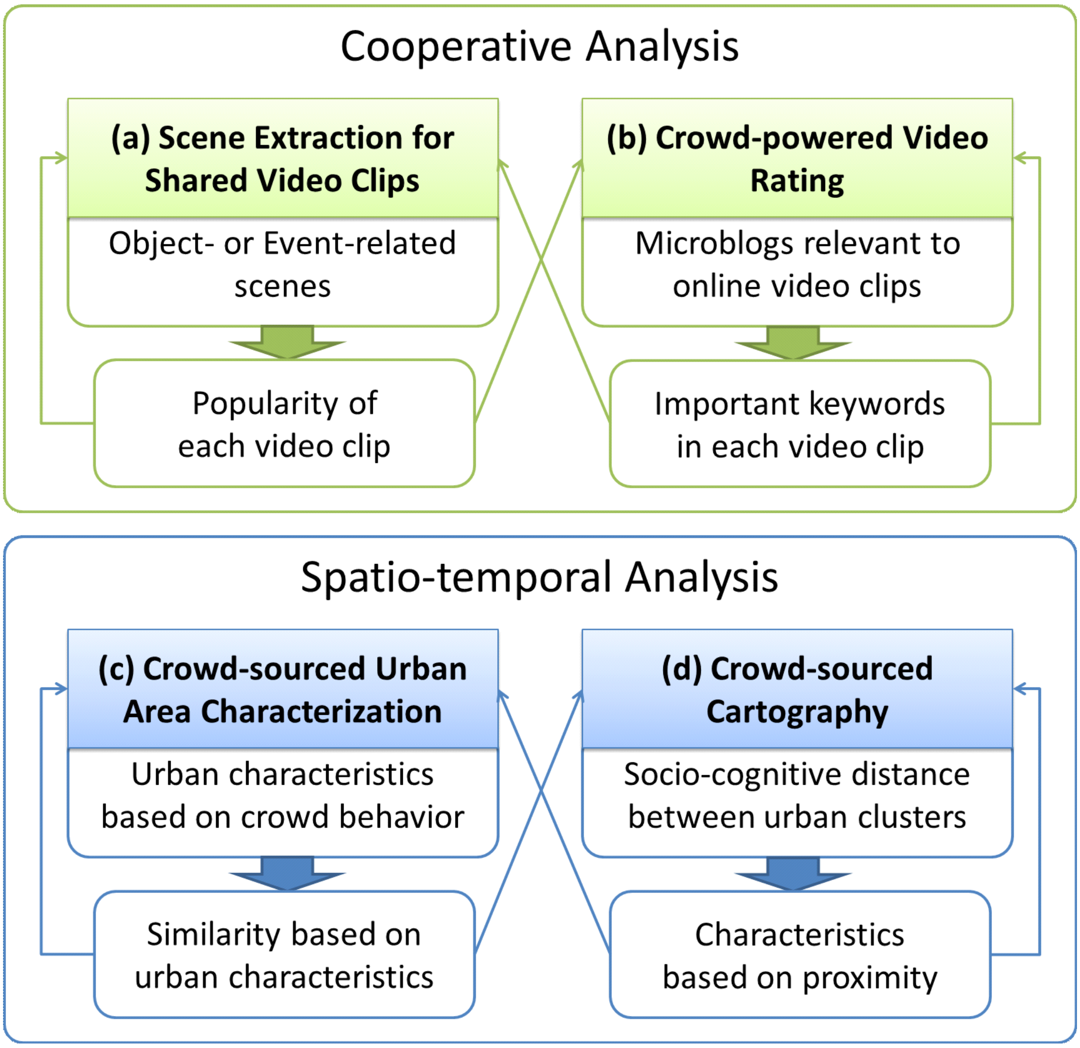
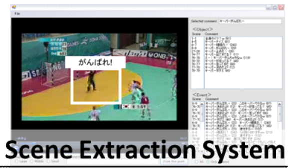
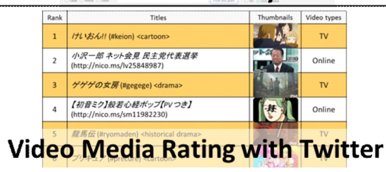
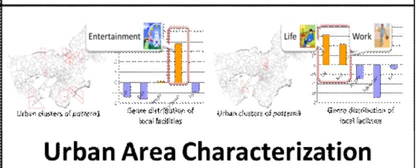
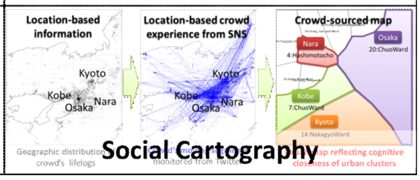


Figure 43: Relations of Our Proposed Approaches and Future Direction

Chapter 8 Conclusions

In this doctoral dissertation, in order to explore advantages of exploitation of massive crowd-sourced data over various social media services, we studied on analysis and utilization of crowd-sourced spatio-temporal contexts from social media. For this, we conducted two types of analyses by focusing on two characteristics of social media; cooperative analysis and spatio-temporal analysis. As a whole, we could confirm the availableness of social media contents and further showed the possibility of creating a virtuous cycle of exploitation of social media contents, which are often misleadingly regarded as very noisy data. We proposed our 4 approaches which are summarized in Table 8 and as follows.

Table 8: Summary of Our Approaches

Theme	Analysis	Focus	Achievement
(a)	Relations of user comments	• Pointing region • Temporal duration	 Scene Extraction System
(b)	Relevance between microblogs and videos	• Tweets from Twitter • TV programs (EPGs) • Online video clips	 Video Media Rating with Twitter
(c)	Significant crowd behavioral patterns	Crowd behavior from Twitter	 Urban Area Characterization
(d)	Socio-cognitive distance between urban areas	Crowd movements from Twitter	 Social Cartography

(a) Crowd-powered Scene Extraction for Shared Video Clips We developed a system to enable users of video sharing websites to easily retrieve video scenes relevant to their interests. The system analyzed both text and non-text aspects of a user's comment and then extracted relevant scenes along with attached comments. The text analysis works in tandem with non-text features, namely, the specified pointing region and temporal duration associated with user comments. In this way, our system could support a better-organized extraction of scenes that have been commented on with a higher degree of relevancy than conventional methods, such as using matching keywords. We described our method and the relation between the scenes and discussed a prototype system. In this case, we could show that there is a possibility to utilize user comments' relations analyzed by exploiting non-textual features of the comments.

(b) **Crowd-powered Video Rating** Due to the advance of many social networking sites, social analytics by aggregating and analyzing crowds' lifelogs are attracting a great deal of attention. In the meantime, there is an interesting trend that people watching TVs are also writing Twitter messages pertaining to their opinions. With the utilization of bigger and broader crowds over Twitter, surveying massive audiences' lifestyles will be an important aspect of exploitation of crowd-sourced data. We established a method for rating and showing integrated ranking of videos from several video media by analyzing massive microblogs collected from Twitter where we could easily find crowd voices relative to video watching. In the experiment, we used 119,575 geo-tagged tweets and 838,636 EPG items (24,841 distinct TV programs) or on-line video clips for measuring relevance between a tweet and a TV program or on-line video clip in terms of textual, spatial, and temporal similarities. Consequently, we could show novel ranking by taking into several video media consideration.

(c) **Crowd-sourced Urban Area Characterization** Recent location-based social networking sites are attractively providing us with a novel capability of monitoring massive crowd lifelogs in the real-world space. In particular, they make it easier to collect publicly shared crowd lifelogs in a large scale of geographic area reflecting the crowd's daily lives and even more characterizing urban space through what they have in minds and how they behave in the space. Particularly, we challenged to analyze urban characteristics in terms of crowd behavior by utilizing crowd lifelogs in urban area over the social networking sites. In order to collect crowd behavioral data, we exploited the most famous microblogging site, Twitter. We first presented a model to deal with crowds' behavioral logs on the social network sites as a representing feature of urban space's characteristics, which was used to conduct crowd-based urban characterization. Based on this crowd behavioral feature, we extracted 4 significant crowd behavioral patterns in a period of time. In the experiment, we collected 1,891,186 geo-tagged tweets from 39,898 users between Jun. 5, 2010 and Jul. 5, 2010 in Kinki area and extracted 4 significant crowd behavioral patterns. Then, we examined relations between the urban clusters grouped based on the patterns and the genres of categories of local facilities.

(d) Crowd-sourced Cartography On behalf of the rapid urbanization, urban areas are gradually becoming a sophisticated space where we often need to know ever evolving features to take the most of the space. Therefore, keeping up with the dynamic change of urban space would be necessary, while it usually requires lots of efforts to understand newly visiting and daily changing living spaces. In order to explore and exploit the urban complexity from crowd-sourced lifelogs, we focused on location-based social network sites. Particularly, we attempted to exploit crowd-sourced location-based lifelogs for generating a socio-cognitive map, whose purpose was to deliver much simplified and intuitive perspective of urban space. For the purpose, we collected $157,097$ geo-tagged tweets published from 25,674 users on April 23, 2012 (a weekday) in Kinki area and measured socio-cognitive distance among urban clusters based on crowd movements. Finally, we could show the possibility to represent socio-cognitive distances between urban clusters and influential strengths of urban clusters on a map.

In future work, we will further study on **Social Big Data Analytics** and **Advanced Utilization**. In order to deal with social media contents generated by numerous number of users in a large-scale geographic area, we will consider how to efficiently manage and analyze ever-increasingly massive crowd-sourced contents in real time. In addition, by analyzing the archived crowd-sourced data over a long duration, we will comprehend dynamics and implications of results. Furthermore, by stepping into analyses for sentiment and human relationships and interactions on social networks, we will develop advanced applications which are useful in real life, such as urban analysis, urban planning, advertisement, and e-commerce.

Acknowledgments

I would like to express my deepest gratitude to my doctoral thesis supervisor, Professor Kazutoshi Sumiya at University of Hyogo, for his valuable advice, helpful discussions, and warm encouragement throughout this study. I would like to appreciate other members of my doctoral dissertation committee, Professor Toru Fukushima at University of Hyogo, Professor Hidetoshi Ikeno at University of Hyogo, and Professor Haruo Yokota at Tokyo Institute of Technology. They took time from their busy schedule to supervise my doctoral thesis by giving valuable comments and advices with their expert knowledge in different fields; systems biology, urban sociology, and computer science.

I wish to express my special thanks to Dr. Ryong Lee at National Institute of Information and Communications Technology for his supervision of this research, valuable comments and advices, constructive discussions, kind supports, and warm encouragement. I also would like to thank Dr. Daisuke Kitayama at Kogakuin University and Professor Setsuko Miyamoto at University of Hyogo for their fruitful comments and suggestions, and kind encouragement. I express my gratitude to the secretary of information media laboratory, Ms. Noriko Fukui, and all other members of the laboratory for their valuable discussions and advices, cooperation, and warm encouragement.

Finally, I would like to thank my parents, Jun Wakamiya and Kumiko Wakamiya, and my adorable dog, Luna Wakamiya, for their always sincere supports and warm encouragement.

This research was supported in part by Grant-in-Aid for JSPS Fellows 24.9154 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the 7th Microsoft Research IJARC Core Project.

References

[1] *YouTube*, http://www.youtube.com/.

[2] *Nico Nico Doga*, http://www.nicovideo.jp/.

[3] *Twitter*, http://twitter.com/.

[4] *Facebook*, http://www.facebook.com/.

[5] *Foursquare*, https://foursquare.com/.

[6] Ja-Hwung, S. Y.-T. Huang, and V. S. Tseng, "Efficient content-based video retrieval by mining temporal patterns," in *Proc. of the 9th International Workshop on Multimedia Data Mining: held in conjunction with the ACM SIGKDD 2008*, aug 2008, pp. 36–42.

[7] A. Karpenko and P. Aarabi, "Tiny videos: Non-parametric content-based video retrieval and recognition," in *Proc. of the Tenth IEEE International Symposium on Multimedia (ISM 2008)*, apr 2008, pp. 619–624.

[8] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video suggestion and discovery for YouTube: Taking random walks through the view graph," in *Proc. of the 17th International World Wide Web Conference (WWW2008)*, apr 2008, pp. 895–904.

[9] D. Yamamoto and K. Nagao, "iVAS: Web-based video annotation system and its applications," in *Proc. of the3rd International Semantic Web Conference*, 2004.

[10] D. Yamamoto, T. Masuda, S. Ohira, and K. Nagao, "Video scene annotation based on web social activities," in *IEEE MultiMedia*, 2008, pp. 22–32.

[11] T. Masuda, D. Yamamoto, S. Ohira, and K. Nagao, "Video scene retrieval using online video annotation," in *Proceedings of the 2007 conference on New frontiers in artificial intelligence*, ser. JSAI'07. Springer-Verlag, 2008, pp. 54–62.

[12] D. Kitayama, N. Oda, and K. Sumiya, "Organizing user comments in a social video sharing system by temporal duration and pointing region," in *Proc. of International Workshop on Information-explosion and Next Generation Search (INGS2008)*, apr 2008, pp. 55–58.

[13] T. Kimura, K. Sumiya, and H. Tanaka, "A video editing support system using users' gazes," in *Proc. of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM2005)*, aug 2005, pp. 149–152.

[14] H. Uehara and K. Yoshida, "Annotating TV drama based on viewer dialogue - analysis of viewers' attention generated on an internet bulletin board -," in *Proc. of IEEE/IPSJ International Symposium on Applications and the Internet (SAINT2005)*, jan 2005, pp. 334–340.

[15] H. Miyamori, S. Nakamura, and K. Tanaka, "Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web," in *Proc. of the 13th annual ACM international conference on Multimedia (ACM Multimedia2005)*, nov 2005, pp. 853–861.

[16] K. Miura, I. Yamada, H. Sumiyoshi, and N. Yagi, "Automatic generation of a multimedia encyclopedia from tv programs by using closed captions and detecting principal video objects," in *Proc. of the Eighth IEEE International Symposium on Multimedia*, dec 2006, pp. 873–880.

[17] M. S. Dao and N. Babaguchi, "Sports event detection using temporal patterns mining and web-casting text," in *Proc. of the 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams (AREA2008)*, vol. 26, oct 2008, pp. 33–40.

[18] S. Wu and Y. Chen, "Mining nonambiguous temporal patterns for interval-based events," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, jun 2007, pp. 742–758.

[19] J. F. Allen, "Maintaining knowledge about temporal intervals," in *Communications of the ACM*, vol. 26, nov 1983, pp. 832–843.

[20] N. Fukino, Q. Ma, K. Sumiya, and K. Tanaka, "Generating football video summery using news article (in japanese)," the 14th Data Engineering Workshop (DEWS2003), Tech. Rep., mar 2003.

[21] S. Pradhan, K. Tajima, and K. Tanaka, "A query model for retrieving relevant intervals within a video stream," in *Proc. of IEEE International Conference on Multimedia Computing and Systems*, vol. 2, jul 1999, pp. 788–792.

[22] X. Wu, M. Takimoto, S. Satoh, and J. Adachi, "Scene duplicate detection based on the pattern of discontinuities in feature point trajectories," in *Proc. of the 16th ACM international conference on Multimedia (MM'08)*, oct 2008, pp. 51–60.

[23] T. Fukuhara, T. Murayama, and T. Nishida, "Analyzing concerns of people using weblog articles and real world temporal data," in *Proc. of WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, may 2005.

[24] N. S. Glance, M. Hurst, and T. Tomokiyo, "Blogpulse: Automated trend discovery for weblogs," in *Proc. of WWW 2004 Workshop on the Weblogging Ecosystem*, may 2004.

[25] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding micro-blogging usage and communities," in *Proc. of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56–65.

[26] D. Zhao and M. B. Rosson, "How and why people twitter: the role that micro-blogging plays in informal communication at work," in *Proc. of the ACM 2009 International Conference on Supporting Group Work (GROUP 2009)*, 2013, pp. 243–252.

[27] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about twitter," in *Proc. of the first workshop on Online social networks*, 2008, pp. 19–24.

[28] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence on twitter: The million follower fallacy," in *Proc. of International AAAI Conference on Weblogs and Social*, 2010.

[29] B. O'Connor, R. Balasubramanyan, B. Routedge, and N. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in *Proc.of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM'10)*, 2010.

[30] N. A. Diakopoulos and D. A. Shamma, "Characterizing debate performance via aggregated twitter sentiment," in *Proc. of the 28th International Conference on Human factors in computing systems*, 2010, pp. 195–198.

[31] R. Sawai, K. Ariyasu, H. Fujisawa, and Y. Kanatsugu, "Tv program

recommendation method using sns based on collaborative filtering (in japanese)," IPSJ SIG Technical Reports, Tech. Rep. 43, 2010.

[32] S. Wakamiya, R. Lee, and K. Sumiya, "Towards better TV viewing rates: exploiting crowd's media life logs over Twitter for TV rating," in *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC '11, 2011, pp. 39:1–39:10.

[33] ——, "Crowd-powered TV viewing rates: measuring relevancy between tweets and TV programs," in *Proceedings of the 2nd International Workshop on Social Networks and Social Media Mining on the Web*, ser. SNSMW '11, 2011, pp. 390–401.

[34] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, ser. CIKM '10, 2010, pp. 759–768.

[35] Sysomos Inc., *Twitter Statistics for 2010: An in-depth report at Twitter's Growth 2010, compared with 2009*, http://www.sysomos.com/insidetwitter/twitter-stats-2010/, 2010.

[36] L. Barkhuus, B. Brown, M. Bell, S. Sherwood, M. Hall, and M. Chalmers, "From awareness to repartee: sharing location within social groups," in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI '08, 2008, pp. 497–506.

[37] J. Hightower, "From position to place," in *Proceedings of the 2003 Workshop on Location-Aware Computing*, ser. part of the 2003 Ubiquitous Computing Conference, October 2003, pp. 10–12.

[38] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with GPS history data," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10, 2010, pp. 1029–1038.

[39] B. De Longueville, R. S. Smith, and G. Luraschi, ""OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires," in *Proceedings of the 2009 International Workshop on Location Based Social Networks*, ser. LBSN

'09, 2009, pp. 73–80.

[40] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10, 2010, pp. 851–860.

[41] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring millions of footprints in location sharing services," in *Proceedings of Fifth International AAAI Conference on Weblogs and Social Media*, ser. ICWSM '11, 2011, pp. 81–88.

[42] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, ser. LBSN '10, 2010, pp. 1–10.

[43] R. Lee, S. Wakamiya, and K. Sumiya, "Discovery of unusual regional social activities using geo-tagged microblogs," *World Wide Web (WWW) Special Issue on Mobile Services on the Web*, vol. 14, no. 4, pp. 321–349, March 2011.

[44] S. Wakamiya, R. Lee, and K. Sumiya, "Crowd-sourced cartography: Measuring socio-cognitive distance for urban areas based on crowd's movement," in *Proc. of the 4th International Workshop on Location-Based Social Networks*, ser. LBSN '12, 2012, pp. 935–942.

[45] ——, "Crowd-sourced urban life monitoring: Urban area characterization based crowd behavioral patterns from Twitter," in *Proc. of the 6th International Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC '12, 2012, pp. 26:1–26:9.

[46] ——, "Exploring reflection of urban society through cyber-physical crowd behavior on location-based social network," in *Proc. of 3rd International Workshop for Social Network and Social Web Mining*, ser. SNSM '12, 2012, pp. 168–179.

[47] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2012,

pp. 133–140.

[48] T. Kurashima, T. Tezuka, and K. Tanaka, “Blog map of experiences: Extracting and geographically mapping visitor experiences from urban blogs,” in *Proc.of the 13th Intl. Conference on Web Information System Engineering*, 2005, pp. 496–503.

[49] F. Grabler, M. Agrawala, R. W. Sumner, and M. Pauly, “Automatic generation of tourist maps,” *ACM Trans. Graph.*, vol. 27, no. 3, pp. 100:1–100:11, Aug 2008.

[50] H. Shu, C. Qi, and G. Edwards, “Computing geographical reality with animated map language,” in *Proc of the 16th Intl. Conference on Artificial Reality and Telexistence-Workshops*, 2006, pp. 52–56.

[51] A. Mislove, S. Lehmann, Y. Y. Ahn, J. P. Onnela, and J. N. Rosenquist, “Understanding the demographics of twitter users,” in *Proc. of the 5th Intl. AAAI Conference on Weblogs and Social Media*, 2011, pp. 133–140.

[52] *Google Video*, http://www.google.co.jp/videohp.

[53] *Synvie*, http://synvie.net/.

[54] E. Y.-T. Shen, H. Lieberman, and G. Davenport, “What’s next?: Emergent storytelling from video collections,” in *Proc. of the 27th international conference on Human factors in computing systems*, apr 2009, pp. 809–818.

[55] Y. Gong, “An accurate and robust method for detecting video shot boundaries,” in *Proc. of IEEE International Conference on Multimedia Computing and Systems (ICMCS’99)*, vol. 1, jul 1999, pp. 850–854.

[56] S. M. M. Tahaghoghi, H. E. Williams, J. A. Thom, and T. Volkmer, “Video cut detection using frame windows,” in *Proc. of the Twenty-eighth Australasian conference on Computer Science (ACSC ’05)*, 2005, pp. 193–199.

[57] S. Zanetti, L. Zelnik-Manor, and P. Perona, “A walk through the web’s video clips,” in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW ’08)*, jun 2008, pp. 1–8.

[58] C. Saraceno and R. Leonardi, “Identification of successive correlated camera shots using audio and video information,” in *Proc. of International Conference on Image Processing (ICIP’97)*, vol. 3, oct 1997, pp. 116–119.

[59] A. Yoshitaka and M. Miyake, "Scene detection by audio-visual features," in *Proc. of IEEE International Conference on Multimedia and Expo (ICME2001)*, vol. 3, aug 2001, pp. 48–51.

[60] L. Braudy and M. Cohen, "FILM THEORY and CRITICISM," in *Oxford University Press*, 2004.

[61] H. Sundaram and S.-F. Chang, "Determining computable scenes in films and their structures using audio-visual memory models," in *Proc. of the eighth ACM international conference on Multimedia*, dec 2000, pp. 95–104.

[62] J. Wang and T.-S. Chua, "A framework for video scene boundary detection," in *Proc. of the tenth ACM international conference on Multimedia*, dec 2008, pp. 243–246.

[63] J. F. Allen, "Maintaining knowledge about temporal intervals," in *Communications of the ACM*, vol. 26, nov 1983, pp. 832–843.

[64] *Mecab*, http://mecab.sourceforge.net/.

[65] H. Oshima, *Slothlib*, http://www.dl.kuis.kyoto-u.ac.jp/slothlib/.

[66] *Nielsen Media Research*, http://nielsen.com/us/en.html.

[67] *twtv.jp*, http://twtv.jp/tbs_twtv.html.

[68] *TV Kingdom: Japan TV Program Guide*, http://tv.so-net.ne.jp/.

[69] J. C. French, A. L. Powell, and E. Schulman, "Applications of approximate word matching in information retrieval," in *Proc. of the 6th International Conference on Information and Knowledge Management (CIKM 1997)*, 1997.

[70] J. Zobel and A. Moffat, "Inverted files for text search engines," *ACM Computing Surveys (CSUR)*, vol. 38, no. 6, 2006.

[71] *Google maps API*, http://code.google.com/intl/ja/apis/maps/.

[72] S. Mann, *wearcam.org*, http://wearcam.org/eastcampusfire.htm.

[73] G. Bell and J. Gemmell, *Total Recall: How the E-Memory Revolution Will Change Everything.* Penguin Group, 2009.

[74] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones : A case study in rome," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 141–151, 2011.

[75] B. Jiang, C. Claramunt, and M. Batty, "An integration of space syntax into GIS for modelling urban spaces," *International Journal of Applied Earth Observation and Geoinformation*, vol. 2, no. 3–4, pp. 161–171, 2000.

[76] P. Cudré-Mauroux and S. Elnikety, "Graph data management systems for new application domains," *PVLDB*, vol. 4, no. 12, pp. 1510–1511, 2011.

[77] A. Iskold, *Social Graph: Concepts and Issues*, http://www.readwriteweb.com/archives/social_graph_concepts_and_issues.php, 2007.

[78] K. Lynch, *The Image of the City*. The MIT Press, 1960.

[79] T. Tezuka, R. Lee, H. Takakura, and Y. Kambayashi, "Models for conceptual geographical prepositions based on web resources," *Journal of Geographical Information and Decision Analysis*, vol. 5, no. 2, pp. 83–94, December 2001.

[80] ——, "Cognitive characterization of geographic objects based on spatial descriptions in web resources," in *Proc. of the Workshop on Spatial Data and Geographic Information Systems (SpaDaGIS)*, March 2003.

[81] *Twitter Open API*, http://apiwiki.twitter.com/Twitter-Search-API-Method%3A-search.

[82] *Google Geocoding API*, http://code.google.com/intl/ja/apis/maps/documentation/geocoding/.

[83] S. Byers and A. Raftery, "Nearest-neighbour clutter removal for estimating features in spatial point processes," *Journal of the American Statistical Association*, vol. 93, pp. 577–584, 1998.

[84] A. Baddeley, "Analysing spatial point patterns in R," *Workshop notes*, vol. 12, no. 6, pp. 1–199, 2008.

[85] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statiscal Soc., Ser. R*, vol. 39, no. 1, pp. 1–38, 1977.

[86] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, March 1982.

[87] M. D. Berg, O. Cheong, M. V. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications (Third edition)*. Springer-Verlag, 2008.

[88] J. X. Yu, Z. Li, and G. Liu, "A data mining proxy approach for efficient frequent itemset mining," *The VLDB Journal*, vol. 17, pp. 947–970, July 2008.

[89] *Yahoo! Local Search Web API*, http://developer.yahoo.co.jp/webapi/map/localsearch/v1/localsearch.html.

[90] *Japan Standard Industry Classification*, http://www.stat.go.jp/english/index/seido/sangyo/san07-2.htm.

[91] *Google Maps*, https://maps.google.com/.

[92] *Facebook Places*, http://www.facebook.com/about/location.

[93] *Cartogram*, http://www.ncgia.ucsb.edu/projects/Cartogram_Central/types.html, 2013.

[94] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, pp. 226–231.

[95] A. M. Andrew, "Another efficient algorithm for convex hulls in two dimensions," *Information Processing Letters*, vol. 9, no. 5, pp. 216–219, 1979.

[96] J. Kruskal and M. Wish, "Multidimensional scaling," in *Sage University Papers on Quantitative Applications in the Social Sciences*, 1978, pp. 07–011.

[97] F. Aurenhammer, "Voronoi diagrams — A survey of a fundamental geometric data structure," *ACM Computing Surveys*, vol. 23, no. 3, pp. 345–405, sep 1991.

Appendix

論文リスト

学術論文（査読あり）

1. Scene Extraction for Video Clips based on the Relation of Text, Pointing Region and Temporal Duration of User Comments
   Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
   In Proc. of 3rd International Workshop on Management and Interaction with Multimodal Information Content (MIMIC 2009), pp. 289–294, Linz, Austria, September 2009
2. Towards Better TV Viewing Rates: Exploiting Crowd's Media Life Logs over Twitter for TV Rating
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2011), pp. 39:1–39:10, Seoul, Korea, February 2011
3. Discovery of Unusual Regional Social Activities using Geo-tagged Microblogs
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   World Wide Web (WWW), Special Issue on Mobile Services on the Web, Volume 14, Number 4, pp. 321-349 March 2011
4. Crowd-powered TV Viewing Rates: Measuring Relevancy between Tweets and TV Programs
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 2nd International Workshop on Social Networks and Social Media Mining on the Web (SNSMW 2011), Lecture Notes in Computer Science, 2011, Volume 6637, pp. 390–401, Hong Kong, April 2011
5. Urban Area Characterization Based on Semantics of Crowd Activities in Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of Fourth International Conference on Geospatial Semantics (GeoS 2011), Lecture Notes in Computer Science, 2011, Volume 6631, pp. 108–123, Brest, France, May 2011

6. Scene Extraction System for Video Clips using Attached Comment Interval and Pointing Region
   Shoko Wakamiya, Daisuke Kitayama, and Kazutoshi Sumiya
   Multimedia Tools and Applications, Volume 54, Number 1, pp. 7–25, August 2011
7. Crowd-based Urban Characterization: Extracting Crowd Behavioral Patterns in Urban Areas from Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN 2011), pp. 77–84, Chicago, Illinois, USA, November 2011 [Best Student Paper Award]
8. Crowd-sourced Urban Life Monitoring: Urban Area Characterization based Crowd Behavioral Patterns from Twitter
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 6th International Conference on Ubiquitous Information Management and Communication (ICUIMC2012), pp. 26:1–26:9, Kuala Lumpur, Malaysia, February 2012
9. Exploring Reflection of Urban Society through Cyber-Physical Crowd Behavior on Location-based Social Network
   Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of 3rd International Workshop for Social Network and Social Web Mining (SNSM2012), pp. 168–179, Busan, Korea, April 2012
10. Tweet 分析による群衆行動を用いた地域特徴抽出
    李 龍，若宮 翔子，角谷 和俊
    情報処理学会論文誌データベース (TOD54), Vol. 5, No. 2, pp. 36–52, 2012 年 6 月.
11. Measuring Crowd-sourced Cognitive Distance between Urban Clusters with Twitter for Socio-cognitive Map Generation
    Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
    In Proc. of the Fourth International Conference on Emerging Databases-Technologies, Applications, and Theory (EDB 2012), pp. 181–192, Seoul, Korea, August 2012 [Best Paper Award]

12. Crowd-sourced Cartography: Measuring Socio-cognitive Distance for Urban Areas based on Crowd's Movement
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
In Proc. of the 4th International Workshop on Location-Based Social Networks (LBSN 2012), pp. 935–942, Pittsburgh, Pennsylvania, USA, September 2012 [Best Student Paper Award]
13. Looking into Socio-cognitive Relations between Urban Areas based on Crowd Movements Monitoring with Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
日本データベース学会論文誌 (DBSJ Journal), Vol. 11, No. 2, pp. 19–24, 2012 年 10 月
14. Urban Area Characterization based on Crowd Behavioral Lifelogs over Twitter
Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
Personal and Ubiquitous Computing: Volume 17, Issue 4, pp. 605–620, April 2013
15. 位置ベース SNS を通した群衆の移動経験に基づく都市空間の近接性分析
若宮 翔子，李 龍，角谷 和俊
情報処理学会論文誌データベース (TOD58), Vol. 6, No. 3, pp. 159–176, 2013 年 6 月
16. Twitter における群衆の経験に基づく近接地域検索システム
若宮 翔子，李 龍，角谷 和俊
日本データベース学会論文誌 (DBSJ Journal), Vol. 12, No. 1, pp. 97–102, 2013 年 6 月

学術報告 (査読あり)

1. Video Scene Retrieval Method Using Attached User Comments Based on Automatic Query Expansion
Shoko Wakamiya and Kazutoshi Sumiya
日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第 3 回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2010), pp. 23–29, 2010 年 8 月.

2. Crowd-based Image of City: Urban Area Characterization through Geographic Regularity over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第3回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2011), pp. 128–137, 2011年8月
3. Social Cognitive Map Generation based on Crowd's Movement over Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
日本データベース学会 第3回ソーシャルコンピューティングシンポジウム講演論文集 (SoC2012), pp. 49–54, 2012年6月
4. Socio-Cognitive Urban Cartography with SNS-based Crowd's Movement
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第4回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2012), 2012年8月

学術報告 (査読なし)

1. 時区間と指定領域に基づくユーザコメントを用いた共有動画コンテンツ視聴システム
若宮 翔子, 北山 大輔, 角谷 和俊
情報処理学会 関西支部支部大会公演論文集, pp. 211–214, 2008年10月
2. 動画共有システムにおけるユーザコメントの領域と時区間を用いたシーン抽出
若宮 翔子, 北山 大輔, 角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第1回データ工学と情報マネジメントに関するフォーラム (DEIM2009) 論文集，E9-3, 2009年3月
3. 投稿ユーザコメントとWebページ文脈を用いた共有動画シーン検索方式
若宮 翔子, 北山 大輔, 角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第2回データ工学と情報マネジメ

ントに関するフォーラム (DEIM Forum 2010) 論文集，F1-5, 2010 年 3 月

4. Web ページ補完のための共有動画に付与されたユーザコメントを用いたシーン抽出手法
若宮 翔子, 北山 大輔, 角谷 和俊
情報処理学会 創立50周年記念 (第72回) 全国大会講演論文集，pp. 775–776, 2010 年 3 月
5. Twitter-based TV Audience Behavior Estimation for Better TV Ratings
若宮 翔子, 李 龍, 角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011) 論文集, A1-4, 2011 年 2 月
6. Urban Area Characterization by Analyzing Crowd Activities in Twitter
Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
日本データベース学会 第 2 回ソーシャルコンピューティングシンポジウム講演論文集 (SoC2011), pp. 27, 2011 年 6 月
7. 位置依存ソーシャルネットワークにおける行動パターン基づく群衆の分類
若宮 翔子, 李 龍, 角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2012) 論文集, D8-3, 2012 年 3 月
8. Twitter における群衆の経験に基づく近接地域検索システム
若宮 翔子，李 龍，角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013) 論文集, A3-3, 2013 年 3 月
[学生プレゼンテーション賞]
9. レシピサイトにおける料理名の修飾表現に着目したネーミングコンセプト抽出
橘 明穂，若宮 翔子，角谷 和俊
電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013) 論文集, P3-5, 2013 年 3 月.

10. 料理名のネーミングコンセプト抽出によるレシピ推薦システム
    橘 明穂，若宮 翔子，難波 英嗣，角谷 和俊
    ARG 第2回 Webインテリジェンスとインタラクション研究会，pp. 30-31, 2013年5月.
11. Analyzing Distortion of Geo-social Proximity using Massive Crowd Moving Logs over Twitter
    Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya
    日本データベース学会 第4回ソーシャルコンピューティングシンポジウム講演論文集 (SoC2013), pp. 23–28, 2013年6月

学会誌

1. 画像メディアと異種メディアの融合
   角谷 和俊，北山 大輔，若宮 翔子
   映像情報メディア学会誌 : 特集 大規模データを活用した映像メディア処理, Vol. 66, No. 11, pp. 896–902, 2012年11月 [ベストオーサー]

投稿中

1. Exploring Geospatial Cognition based on Location-based Social Network Sites
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   World Wide Web (WWW), Special Issue on Internet and Web Information Systems, 2013
2. Exploring Latent Relations between Urban Areas and Crowd Behavior from Location-based Social Networks for Urban Area Characterization
   Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya
   Journal of Database Management, 2013
3. Crowd Footprints Network for Social-Urban Analytics with Location-based Social Networks
   Ryong Lee, Shoko Wakamiya, Kazutoshi Sumiya
   Distributed and Parallel Databases, Special issue on Data Management and Analysis in Location-Based Social Networks, 2013
4. 画像メディアと異種マルチメディアの融合
   北山 大輔，若宮 翔子，王 元元，角谷 和俊
   画像ラボ, 2013